现代数学基础丛书 162

现代测量误差模型

李高荣　张　君　冯三营　著



科学出版社

www.sciencep.com



科学数理分社
电　话：(010) 64010624
E-mail：wangliping@mail.sciencep.com

销售分类建议：高等数学

定　价：138.00 元

"十二五"国家重点图书出版规划项目

现代数学基础丛书 162

现代测量误差模型

李高荣　张　君　冯三营　著

科 学 出 版 社
北 京

内 容 简 介

现代测量误差模型是统计学热门研究课题之一，并在生物学、医学、传染病学、森林、经济学、金融学和遥感等领域有着广泛的应用. 本书主要研究两种测量误差类型：可加测量误差和扭曲测量误差. 全书共分 8 章内容，主要阐述这两种测量误差的线性模型、非线性模型、部分线性模型、变系数模型、单指标模型和部分线性单指标模型、降维模型的估计和统计推断方法，除了介绍这些模型的发展动态，也详细介绍相关研究的一些最新的研究成果，使读者对测量误差模型的方法和统计思想有一个较为全面的了解，并起到抛砖引玉的作用.

本书主要适用于理工科院校数理统计、数学和计量经济学专业的研究生、高年级大学生、教师、一般科学技术人员阅读；另外，本书可供各行各业应用统计科学工作者和医学、生物学、经济学、金融学、社会学、心理学和工业工程等专业人士参考.

图书在版编目(CIP)数据

现代测量误差模型/李高荣，张君，冯三营著. —北京：科学出版社，2016.3
(现代数学基础丛书)
ISBN 978-7-03-047868-9
I. ①现… II. ①李… ②张… ③冯… III. ①测量误差 IV. ①P207

中国版本图书馆 CIP 数据核字(2016) 第 058800 号

责任编辑：王丽平／责任校对：彭 涛
责任印制：张 伟／封面设计：陈 敬

科学出版社 出版
北京东黄城根北街 16 号
邮政编码：100717
http://www.sciencep.com

北京中石油彩色印刷有限责任公司 印刷
科学出版社发行 各地新华书店经销
*
2016 年 3 月第 一 版 开本：720 × 1000 1/16
2016 年 3 月第一次印刷 印张：23 1/2
字数：440 000

定价：138.00 元

(如有印装质量问题，我社负责调换)

作 者 简 介

李高荣 北京工业大学统计学教授, 博士生导师, 2007 年于北京工业大学获博士学位, 2007 年 8 月到 2009 年 6 月在华东师范大学金融与统计学院从事博士后研究工作, 2009 年 7 月到北京工业大学工作. 目前担任全国工业统计学教学研究会常务理事、中国现场统计研究会高维数据统计分会理事、生存分析分会副秘书长、北京应用统计学会常务理事、美国数学评论评论员及众多国内外统计学术期刊的审稿专家. 主要研究方向是非参数统计、经验似然、变量选择、复杂高维数据分析等. 迄今为止, 在科学出版社出版过专著《纵向数据半参数模型》, 发表学术论文 60 余篇, 其中有 40 多篇被 SCI 和 EI 收录. 2010 年入选北京市属高等学校人才强教深化计划"中青年骨干人才培养计划"和 "北京市优秀人才培养资助计划", 2012 年破格评为北京工业大学"京华人才".

张君 深圳大学统计学讲师, 硕士生导师, 香港浸会大学和深圳大学联合培养博士生项目合作导师, 2012 年于华东师范大学获博士学位, 2012 年 9 月到深圳大学数学与统计学院工作, 现任深圳大学统计科学研究所所长助理. 目前担任广东省现场统计学会理事、中国现场统计研究会高维数据统计分会理事、美国数学评论评论员及众多国内外统计学术期刊的审稿专家. 主要研究方向是非参数统计、降维、变量选择和测量误差数据分析等. 主持和参与了国家自然科学基金、天元基金等国家和省部级科研项目 7 项, 发表 SCI 学术论文 20 多篇. 2014 年入选深圳市海外高层次人才"孔雀计划"(C 类), 2015 年入选深圳大学优秀青年教师培养计划.

冯三营 郑州大学数学与统计学院讲师, 硕士生导师, 美国数学评论评论员. 2004 年毕业于郑州大学获学士学位, 2007 年毕业于北京工业大学获硕士学位, 2015 年毕业于北京工业大学获统计学博士学位. 2015 年获北京市优秀毕业生和北京工业大学"十佳毕业生"称号, 荣获北京工业大学优秀博士学位论文. 主要研究方向是非参数统计、经验似然、测量误差、变量选择和函数型数据分析等. 主持和参与了国家自然科学基金、国家社科基金等国家和省部级科研项目 8 项, 在国内外学术期刊发表学术论文 30 余篇, 其中 15 篇被 SCI 和 SSCI 收录.

《现代数学基础丛书》序

对于数学研究与培养青年数学人才而言，书籍与期刊起着特殊重要的作用．许多成就卓越的数学家在青年时代都曾钻研或参考过一些优秀书籍，从中汲取营养，获得教益．

20世纪70年代后期，我国的数学研究与数学书刊的出版由于文化大革命的浩劫已经破坏与中断了 10 余年，而在这期间国际上数学研究却在迅猛地发展着．1978 年以后，我国青年学子重新获得了学习、钻研与深造的机会．当时他们的参考书籍大多还是50年代甚至更早期的著述．据此，科学出版社陆续推出了多套数学丛书，其中《纯粹数学与应用数学专著》丛书与《现代数学基础丛书》更为突出，前者出版约 40 卷，后者则逾 80 卷．它们质量甚高，影响颇大，对我国数学研究、交流与人才培养发挥了显著效用．

《现代数学基础丛书》的宗旨是面向大学数学专业的高年级学生、研究生以及青年学者，针对一些重要的数学领域与研究方向，作较系统的介绍．既注意该领域的基础知识，又反映其新发展，力求深入浅出，简明扼要，注重创新．

近年来，数学在各门科学、高新技术、经济、管理等方面取得了更加广泛与深入的应用，还形成了一些交叉学科．我们希望这套丛书的内容由基础数学拓展到应用数学、计算数学以及数学交叉学科的各个领域．

这套丛书得到了许多数学家长期的大力支持，编辑人员也为其付出了艰辛的劳动．它获得了广大读者的喜爱．我们诚挚地希望大家更加关心与支持它的发展，使它越办越好，为我国数学研究与教育水平的进一步提高做出贡献．

杨　乐

2003 年 8 月

序　一

在工程、生物、医学、传染病学、经济、金融、社会和遥感等各个领域中, 人们不管用何种方式收集数据, 不可避免会收集到测量误差数据, 但是在进行数据分析时往往会忽略测量误差的影响, 导致分析结果不准确或存在较大的偏差. 19 世纪 70 年代, 统计学家对测量误差数据分析提出线性测量误差模型以来, 测量误差模型的研究已经取得了非常丰富的研究成果. 但是目前国内外有关测量误差模型的专著主要是以参数测量误差模型为主, 而对于非参数或半参数测量误差模型的研究, 目前还缺乏这方面的总结性工作, 特别是近几年提出的扭曲测量误差的协变量调整模型.《现代测量误差模型》这本专著以可加测量误差数据和扭曲测量误差数据为背景, 系统介绍在这两种测量误差数据下测量误差模型的估计和统计推断方法, 以及理论结果.

该书第一作者李高荣教授是具有一定学术造诣的青年统计学者, 近十年来一直致力于非参数统计和数据分析方面的研究, 取得了一系列创新性成果. 第二作者张君博士一直从事具有扭曲测量误差数据的协变量调整模型的研究, 在该领域取得了可喜的研究成果. 作者以自己多年的研究工作为基础, 并参考国内外的有关文献和新进展, 系统地总结测量误差模型研究的思想、方法和理论结果, 集中体现近年来在这方面的研究成果和所形成的理论体系, 丰富测量误差模型的研究内容, 具有重要的理论和应用价值. 该书有以下特色:

(1) 内容取材先进, 反映了国内外测量误差模型研究的最新成果, 重点突出了测量误差模型研究的思想和方法, 以及理论结果;

(2) 撰写和阐述具有科学性, 理论性强, 结构清晰严谨, 思想、方法、算法、定理、模拟研究和实例分析完整充实;

(3) 著作具有广泛的适应性, 不仅适用于理工科院校数理统计、数学和计量经济学专业的学生和教师, 还适用于数据分析的实际工作者;

(4) 著作撰写规范、文笔流畅、言简意赅、图表清晰.

因此, 该著作具有很高的学术水平, 是一部优秀的学术专著, 将为促进统计学的发展发挥重要作用.

王启华

中国科学院数学与系统科学研究院

2015 年 10 月 18 日

序　二

在一年多前就听李高荣教授讲他正在组织撰写一本关于半参数测量误差模型的专著, 当时就很期待这部专著的出版. 首先, 国内目前还没有一本专门介绍测量误差模型的专著, 其次, 测量误差是实际问题中常见的一种数据类型, 在工程、生物、医学、经济和金融等各个领域普遍存在, 而对测量误差模型的研究也一直是统计学科研人员和数据分析工作者特别关心的问题. 现在由衷地高兴, 李高荣、张君和冯三营高质量地完成了《现代测量误差模型》这部专著, 倾注了他们大量的辛勤劳动和心血.

全书共有 8 章内容, 介绍了可加测量误差数据和扭曲测量误差数据下线性测量误差模型、非线性测量误差模型、部分线性测量误差模型、变系数测量误差模型、单指标测量误差模型、协变量调整模型和测量误差降维模型的估计和统计推断方法. 该书的一个显著优点是收集材料多, 内容丰富, 结构合理, 除了介绍传统可加测量误差模型的统计思想和方法外, 还系统介绍了 2005 年发展起来的协变量调整模型的方法和理论, 以及测量误差模型的估计、假设检验、变量选择和降维等方法. 尽管该书的内容较广, 但是作者对每种方法的阐述却是重点突出, 系统性强. 作者按照模型的类型这一主线介绍统计思想、方法和理论, 这又是该书的一大优点. 该书除了注重统计思想、方法和理论, 更加重视在统计学上的应用, 以及实际数据的分析, 这是本书的一个特色. 因此, 该书既适用于数理统计的教师和研究生, 也适用于从事数据分析的工作者.

该书三位作者一直从事非参数统计与数据分析的科研工作, 对测量误差模型的研究取得了创新性的成果. 该书以他们多年的研究工作为基础, 凝练相关文献的精华和最新进展, 突出了该书的特色和优点, 相信本书会成为一本相当有参考价值的参考读物. 更为重要的是, 该书系统总结了近几年测量误差模型所取得的研究成果, 具有一定的学术价值.

三位青年学者能够静下心来, 埋头完成这样一本学术专著, 实属不易. 感谢他们对统计学科发展做出的贡献.

薛留根
北京工业大学应用数理学院
2015 年 10 月 20 日

前　言

测量误差数据不可避免地存在于实际问题的研究中, 而把带有测量误差数据所建立的模型称为测量误差回归模型, 该类模型存在于许多领域, 如生物学、医学、传染病学、森林、经济学、金融学和遥感等领域. 如果忽略测量误差的影响, 将导致统计推断结果不准确或存在一定的偏差. 但是统计学者和实际工作者在对测量误差数据进行统计推断和分析时, 仍然希望能够正确分析变量之间的关系, 从而进行有效的统计推断和统计预测等. 因此, 越来越多的统计学家和实际工作者认识到处理测量误差的重要性和必要性. 在测量误差模型方面已有不少文献讨论, 如 Fuller (1987) 对线性测量误差模型的研究做了概括和总结, Carroll 等 (1995) 对非线性测量误差模型的研究做了详细的讨论. 本书目的是系统介绍近几年半参数测量误差模型的研究方法和理论结果, 使读者对测量误差领域有更深入的了解.

本书主要考虑两种测量误差数据: 可加测量误差数据和扭曲测量误差数据, 介绍在这两种测量误差数据下线性测量误差模型、非线性测量误差模型、部分线性测量误差模型、变系数测量误差模型、单指标测量误差模型、协变量调整模型和测量误差降维模型的估计和统计推断方法. 本书除了介绍这些测量误差模型的发展动态和一些已取得的研究成果外, 也详细介绍了有关它们研究的一些最新的研究成果, 让读者对测量误差模型的方法和统计思想有一个较为全面的了解和认识. 本书的一个特点是只介绍各种测量误差模型的统计方法和统计思想, 书中给出的理论结果没有给详细的证明过程, 感兴趣的读者可参考相关的参考文献. 本书的每章内容自成体系, 读者可选择感兴趣的章节阅读.

全书分为 8 章. 第 1 章绪论, 主要介绍一些预备知识, 主要内容包括: 回归模型介绍, 非参数光滑方法, 经验似然方法, bootstrap 方法, 变量选择方法等. 此外, 以一元线性测量误差模型和非参数测量误差模型为例, 介绍一些本书中涉及的估计方法, 让读者了解研究测量误差模型的必要性和重要性. 最后对扭曲测量误差数据和协变量调整回归模型做了简单介绍. 第 2 章针对线性可加测量误差模型和协变量调整线性回归模型, 讨论两种测量误差模型的估计方法和理论结果. 第 3 章考虑非线性测量误差模型和非线性半参数测量误差模型的统计推断问题, 介绍这两种模型的估计方法和理论结果. 第 4 章讨论多元混淆变量情形下的协变量调整线性模型和非线性模型, 为了避免 “维数祸根” 问题, 本章把多元混淆变量的失真函数建模为单指标模型或可加模型, 发展估计方法并研究模型中参数的估计和统计推断问题, 建立估计量和检验统计量的渐近性质. 第 5 章讨论两种测量误差的部分线性模型:

部分线性 EV 模型和协变量调整部分线性模型. 研究这两种模型的估计方法, 经验似然和变量选择, 以及所得估计量的渐近结果. 第 6 章考虑变系数测量误差模型和部分线性变系数测量误差模型的估计方法、变量选择和统计推断问题, 并研究估计量和检验统计量的渐近性质. 第 7 章针对单指标模型和部分线性单指标模型, 在可加测量误差和受失真函数污染后的测量误差两种情况下, 讨论模型中参数和非参数分量的估计方法和理论结果. 第 8 章首先介绍几种经典的充分降维方法, 并讨论结构维数的选取问题. 然后针对可加测量误差降维模型和协变量调整降维模型讨论了充分降维方法及其相关理论.

本书第 1 章、第 2 章、第 5 章 (除 5.6 节的内容)、第 6 章 6.5.2 节和第 7 章 7.2 节由李高荣负责撰写, 第 4 章、第 5 章 5.6 节、第 7 章 (除 7.2 节的内容) 和第 8 章由深圳大学数学与统计学院张君负责撰写, 第 3 章和第 6 章 (除 6.5.2 节的内容) 由郑州大学数学与统计学院冯三营负责撰写. 李高荣制定了全书的结构和框架, 对全书的内容进行了认真的修改和校对, 最后对全书进行了定稿.

本书的完成得到了国家自然科学基金 (11471029, 11401391, 11501522)、北京市自然科学基金 (1142002), 北京市教育委员会科技计划面上项目 (KM201410005010) 和北京工业大学北京科学与工程计算研究院开放基金资助项目的支持, 还特别得到了北京工业大学京华人才支持计划项目 (2013-JH-L07) 对本书的资助, 作者谨在此表示感谢.

感谢我的家人和爱女李睿菡这些年对我科研工作的支持, 让我有足够的时间和精力完成本书的部分撰写工作和全书的修改. 特别感谢张君和冯三营, 在我的要求下参与本书的撰写工作. 他们的加入使得本书的内容更加完善, 并保证本书能够按时顺利完成, 同时也感谢他们家人对他们科研工作的支持. 最后感谢我的导师薛留根教授多年来对我的关心和支持, 也感谢北京工业大学应用数理学院和北京科学与工程计算研究院支持和帮助我的各位同事. 在本书的出版过程中, 得到了科学出版社领导和王丽平等编辑的支持和帮助, 在此一并表示感谢.

由于作者知识和能力有限, 书中内容难免有许多不足之处, 还恳请各位专家和同行及实际应用者多提宝贵意见, 敬请有关专家与广大读者批评指正.

李高荣

北京工业大学北京科学与工程计算研究院

2015 年 10 月

符 号 表

$\mathbb{R}$	实数集合
$\mathbb{R}^p$	p 维 Euclidean 空间
$\|\cdot\|$	Euclidean 模
i.i.d.	独立同分布
$=:$	“定义为” 或 “记为”
T	向量或矩阵的转置
$\mathrm{tr}(A)$	矩阵 A 的迹
$A \otimes B$	矩阵 A 和 B 的 Kronecker 乘积
$A^{\otimes 2}$	AA^{T}
$\mathrm{mineig}(A)$	矩阵 A 的最小特征值
$\mathrm{diag}(a_1, \cdots, a_n)$	由元素 $a_1, \cdots, a_n$ 组成的对角矩阵
$A_{(i,j)}$	矩阵 A 的第 (i,j) 个位置上的元素
$\xrightarrow{L}$	依分布收敛
$\xrightarrow{P}$	依概率收敛
$a.s.$	几乎处处收敛, 或者依概率 1 强收敛
$y = O(1)$	y 是有界变量
$y = o(1)$	y 是无穷小量
$\xi_n = o_P(\eta_n)$	对任一 $\epsilon > 0$, 有 $P\{\|\xi_n\| \geqslant \epsilon \|\eta_n\|\} \to 0$
$\xi_n = o_P(1)$	ξ_n 依概率收敛到 0
$\xi_n = O_P(1)$	ξ_n 依概率有界
$K(\cdot)$	核函数
h和h_n	窗宽
$N(\mu, \Sigma)$	均值为 μ, 协方差阵为 Σ 的正态分布
χ_p^2	自由度为 p 的 χ^2 分布
$t(n)$	自由度为 n 的 t 分布
$\|\cdot\|_p$	p 范数
$\mathrm{sgn}(t) = I(t > 0) - I(t < 0)$	符号函数
$\perp\!\!\!\perp$	表示独立
$\|g\|_\infty$	对于任意的函数 $g(u)$, 定义 $\|g\|_\infty = \sup\limits_{u \in [0,1]} \|g(u)\|$
$\|A\|_{\mathrm{F}}$	表示矩阵 A 的 Frobenius 范数

目　录

第 1 章　绪　　论

为了方便读者阅读本书后面各个章节的内容, 本章首先介绍一些预备知识, 主要内容包括: 本书所涉及的回归模型介绍、非参数光滑方法、经验似然方法、bootstrap 方法、基于惩罚函数的变量选择方法; 其次, 以一元线性测量误差模型和非参数测量误差模型为例, 介绍一些本书中涉及的估计方法, 让读者了解研究测量误差模型的必要性和重要性; 最后, 对受失真函数污染的扭曲测量误差数据和协变量调整回归模型做了简单介绍.

1.1　回归模型介绍

回归模型主要是用于研究响应变量和协变量之间的关系, 并进行定量描述的一种数学模型. 回归模型可以用于控制和预测, 并给出这种控制或预测的精确程度. 自从 Galton 于 1886 年提出回归模型的概念, 从经典的线性回归模型发展至今, 在解决实际问题中, 实际工作者和学者们不断发展和提出了许多类型的回归模型, 例如广义线性模型、部分线性模型、变系数模型和单指标模型等. 下面就本书中涉及的几种回归模型进行简要介绍.

1.1.1　参数回归模型

令 Y 是响应变量, $X=(X_1,\cdots,X_p)^{\mathrm{T}}$ 是 p 维的协变量向量. 给定协变量 X 的条件下, 参数回归模型可以表示为如下的形式:

$$E(Y|X)=f(X,\beta), \tag{1.1.1}$$

其中 $f(\cdot,\cdot)$ 表示已知函数, $\beta=(\beta_1,\cdots,\beta_q)^{\mathrm{T}}$ 为 q 维的未知参数向量.

模型 (1.1.1) 包含经典的线性模型、广义线性模型和非线性回归模型. 如果 $f(X,\beta)=\beta_0+X_1\beta_1+\cdots+X_p\beta_p$, 则有如下经典的线性模型:

$$Y=\beta_0+X_1\beta_1+\cdots+X_p\beta_p+\varepsilon,$$

其中 ε 为模型误差, 通常假设模型误差是均值为 0 的随机误差. 如果响应变量 Y 为非高斯分布, 如泊松 (Poisson) 分布、伽马 (Gamma) 分布、二项 (Binomial) 分布、指数 (Exponential) 分布等时, 则有下面的广义线性模型:

$$E(Y|X)=f(X^{\mathrm{T}}\beta), \tag{1.1.2}$$

其中 $f(\cdot)$ 为已知的函数, β 为 p 维的未知回归参数向量. 通常把 $h(\cdot) = f^{-1}(\cdot)$ 称为已知的联系函数 (link function). 关于广义线性模型的详细讨论见 McCullagh 和 Nelder (1990). 而对于一般的已知函数 $f(\cdot,\cdot)$, 即考虑如下的模型:

$$Y = f(X, \beta) + \varepsilon, \tag{1.1.3}$$

其中 ε 为均值为 0 的模型误差. 通常把模型 (1.1.3) 称为非线性回归模型, 可以采用非线性最小二乘方法获得模型中未知参数向量 β 的估计.

1.1.2　非参数模型

假设 Y 为响应变量, X 为影响 Y 的协变量, 则非参数回归模型的形式为:

$$Y = g(X) + \varepsilon, \tag{1.1.4}$$

其中 $g(x) = E(Y|X = x)$ 为未知的回归函数, ε 为模型误差, 且满足 $E(\varepsilon|X) = 0$.

非参数回归模型的优点是回归函数 $g(\cdot)$ 的形式任意, 而且模型的假设少, 可以很好地拟合实际数据. 但非参数回归模型的缺点是当 $X \in \mathbb{R}^p$, 且 X 的维数 p 较高时, 对非参数模型进行估计和统计推断时, 会遇到所谓的 "维数祸根" 问题. 在实际应用中, 为了保留参数模型的优点以及非参数回归模型数据适应性的优点, 同时避免 "维数祸根" 问题, 统计学者提出并发展了很多半参数回归模型. 例如, 部分线性模型、变系数模型、单指标模型、部分线性单指标模型等, 这些模型已经被广泛应用到了生物医学和计量经济学等领域中.

1.1.3　半参数模型

考虑如下的半参数模型:

$$Y = f(X, \beta) + g(T) + \varepsilon, \tag{1.1.5}$$

其中 $X = (X_1, \cdots, X_p)^{\mathrm{T}}$ 为 p 维的协变量向量, $\beta = (\beta_1, \cdots, \beta_q)^{\mathrm{T}}$ 是 q 维的未知参数向量, $f(\cdot,\cdot)$ 是已知的函数, $g(\cdot)$ 是一元未知的光滑函数, ε 是期望为 0 的随机模型误差. 模型 (1.1.5) 称为非线性半参数模型, 其中响应变量 Y 以非线性函数 $f(\cdot,\cdot)$ 依赖于 p 维协变量 X 和 q 维参数向量 β, 而 Y 与 T 之间呈非线性关系.

当 $q \equiv p$, 且 $f(X, \beta) = X^{\mathrm{T}}\beta$ 时, 模型 (1.1.5) 退化到一个特别重要的半参数模型, 即部分线性模型, 模型的形式为:

$$Y = X^{\mathrm{T}}\beta + g(T) + \varepsilon. \tag{1.1.6}$$

自 Engle 等 (1986) 在研究气象条件对电力需求影响这一实际问题首次提出部分线性模型 (1.1.6) 以来, 该模型已出现了一系列丰富的研究成果, 其研究也不断向各个

方向发展, 例如对各种复杂数据, 如在纵向数据、面板数据、删失数据、缺失数据、测量误差数据和相依混合序列数据等情形下的研究也都取得了可喜的研究成果. 详细的成果介绍见柴根象和洪圣岩 (1995), Härdle 等 (2000), 薛留根 (2012a), 李高荣和杨宜平 (2015) 等.

1.1.4 变系数模型

在实际应用中, 统计学家结合经典线性回归模型容易解释的优点, 半参数模型的灵活性, 以及平衡建模偏差和高维数据的 "维数祸根" 问题等优点, 提出了部分线性变系数模型, 具体形式为:

$$Y = X^{\mathrm{T}}\beta + Z^{\mathrm{T}}\alpha(T) + \varepsilon, \tag{1.1.7}$$

其中 $\alpha(\cdot) = (\alpha_1(\cdot), \cdots, \alpha_q(\cdot))^{\mathrm{T}}$ 是一个 q 维的未知回归系数函数, $\beta = (\beta_1, \cdots, \beta_p)^{\mathrm{T}}$ 是一个 p 维的未知回归系数向量, Y 是响应变量, $(X^{\mathrm{T}}, Z^{\mathrm{T}}, T)^{\mathrm{T}}$ 是相关的协变量, ε 是一个独立的随机误差, 且几乎处处有 $E(\varepsilon|X, Z, T) = 0$.

明显地, 模型 (1.1.7) 是经典线性模型和半参数模型的推广, 它包含了许多重要的统计模型. 例如, 当 $\alpha(\cdot) \equiv \theta$ 时, 其中 θ 是一个参数向量, 模型 (1.1.7) 便成为经典的线性回归模型. 当 $q = 1$ 且 $Z = 1$ 时, 模型 (1.1.7) 退化成了部分线性回归模型 (1.1.6). 当 $X = 0$ 时, 模型 (1.1.7) 变成了如下著名的变系数模型:

$$Y = Z^{\mathrm{T}}\alpha(T) + \varepsilon. \tag{1.1.8}$$

对于模型 (1.1.8), 当 $Z = (1, \cdots, 1)^{\mathrm{T}}$ 时, 则变系数模型 (1.1.8) 退化为非参数可加模型 (Hastie and Tibshirani, 1990). 模型 (1.1.7) 和模型 (1.1.8) 是非参数模型和半参数模型的推广形式, 因此该类模型有较强的适应性, 并且有着广泛的应用背景. 目前已有很多统计学者对变系数模型进行了研究, 如 Hastie 和 Tibshirani (1993), Fan 和 Zhang (2008), 张日权和卢一强 (2004).

1.1.5 单指标模型

部分线性单指标模型考虑响应变量 Y 可能与 p 维解释变量 X 之间是线性关系, 同时与 q 维解释变量 Z 之间是非线性关系, 即把 q 维解释变量 Z 投影到一个线性空间上, 然后在这个一维线性空间上拟合一个一元函数. 部分线性单指标模型的形式如下:

$$Y = X^{\mathrm{T}}\beta + g(Z^{\mathrm{T}}\theta) + \varepsilon, \tag{1.1.9}$$

其中 β 为 p 维未知参数向量, 刻画了 Y 与 X 之间的线性关系, θ 为 q 维未知参数向量, 刻画了 Z 的线性组合, $g(\cdot)$ 为一元未知联系函数, 函数 $g(\cdot)$ 和线性组合 $Z^{\mathrm{T}}\theta$ 刻画了 Y 与 Z 之间的非线性关系, ε 为均值为 0 的随机误差, 且有 $E(\varepsilon|X, Z) = 0$.

为了模型的可识别性, 需要假定 $\|\theta\|=1$, 且 θ 的第一个非零元素大于 0, 其中 $\|\cdot\|$ 表示 Euclidean 模.

部分线性单指标模型 (1.1.9) 是一个比较广泛的模型, 包含许多重要的统计模型. 如:

(1) 若 θ 为一维变量, 则模型 (1.1.9) 就退化为部分线性模型;

(2) 若 $X=0$ 且 θ 为一维变量, 则模型 (1.1.9) 就简化为一元非参数模型;

(3) 若 $X=0$, $g(\cdot)$ 为正态分布函数或 Logistic 分布函数时, 则模型 (1.1.9) 就成为 Probit 模型或 Logistic 模型;

(4) 若 $g(\cdot)=0$, 则模型 (1.1.9) 就成为经典的线性模型;

(5) 若 $X=0$, 则模型 (1.1.9) 就退化为如下的单指标模型:

$$Y=g(Z^{\mathrm{T}}\theta)+\varepsilon. \tag{1.1.10}$$

单指标模型作为一种广义的回归模型, 是 20 世纪 80 年代中后期发展起来的一种重要的统计模型, 该模型对解释变量 X 使用降维方法, 避免了“维数祸根”问题, 同时又把参数模型和非参数模型有机地结合起来, 对实际数据有很好的拟合效果和更大的适用性. 因此, 单指标模型不仅有实际的应用背景, 而且有广泛的应用前景, 在金融经济、计量经济学和生物数学等领域具有很大的应用价值. 该模型的研究已经成为目前统计界的热门课题之一. 关于单指标模型的研究现状介绍, 见薛留根 (2012b, 2012c).

1.2　非参数光滑方法

本节针对非参数回归模型 (1.1.4), 介绍三种非参数光滑方法: N-W 核光滑方法, Gasser-Müller 光滑方法和局部多项式光滑方法, 并介绍非参数光滑方法中光滑参数的选取方法.

1.2.1　N-W 核光滑方法

1. N-W 核估计

假设 $\{(Y_i,X_i),i=1,\cdots,n\}$ 是来自模型 (1.1.4) 的 i.i.d.的随机样本, 另外为了简单, 假设协变量 X_i 的支撑集是 [0,1]. 令 $K(\cdot)$ 是一个核函数, h 为窗宽, 主要用于控制局部区域的大小. Nadaraya (1964) 与 Watson (1964) 分别提出了回归函数 $g(\cdot)$ 的核光滑 (kernel smoothing) 估计方法, 其主要思想是: 对于变量 X_i 支撑集内任意给定的一点 x, 在 x 的一个邻域 $[x-h,x+h]$ 中, 假定回归函数 $g(\cdot)$ 为一个常数, 如果 $h\to 0$, 这时在局部邻域 $[x-h,x+h]$ 内, 极小化加权最小二乘目标函数,

则可得到 Nadaraya-Watson 核估计, 简称为 N-W 核估计, 定义为

$$\begin{aligned}\hat{g}_{\mathrm{NW}}(x) &= \arg\min_{\theta}\sum_{i=1}^{n}(Y_i-\theta)^2 K\left(\frac{X_i-x}{h}\right)\\ &= \sum_{i=1}^{n} W_{ni}(x) Y_i,\end{aligned} \tag{1.2.1}$$

其中

$$W_{ni}(x) = \frac{K\left(\dfrac{X_i-x}{h}\right)}{\displaystyle\sum_{j=1}^{n} K\left(\dfrac{X_j-x}{h}\right)}$$

为核权函数, $i=1,\cdots,n$. 从式 (1.2.1) 容易看出, N-W 核估计 $\hat{g}_{\mathrm{NW}}(x)$ 是响应变量 Y_i 的加权平均值. 图 1.2.1 给出了 N-W 核估计的原理示意图.

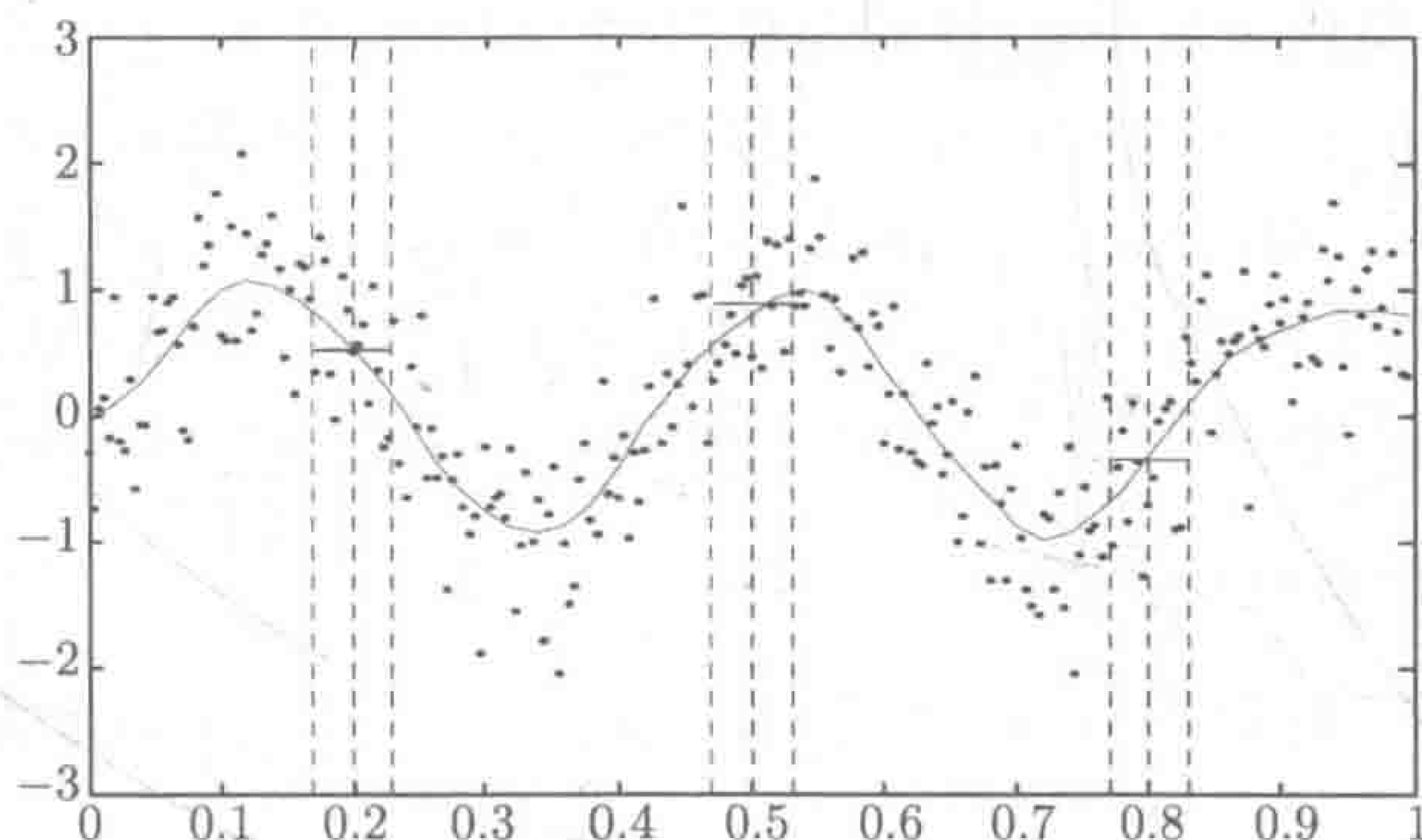


图 1.2.1 N-W 核估计的原理示意图

另外, 核函数 $K(\cdot)$ 通常取为某个概率密度函数, 常用的核函数有:

(1) 均匀核: $K(u)=\dfrac{1}{2}I(|u|\leqslant 1)$;

(2) 高斯核: $K(u)=\dfrac{1}{\sqrt{2\pi}}\exp(-u^2/2)$;

(3) Epanechnikov 核: $K(u)=\dfrac{3}{4}(1-u^2)I(|u|\leqslant 1)$;

(4) 四次 (quartic) 核: $K(u)=\dfrac{15}{16}(1-u^2)^2 I(|u|\leqslant 1)$.

图 1.2.2 给了上面四个核函数的曲线图.

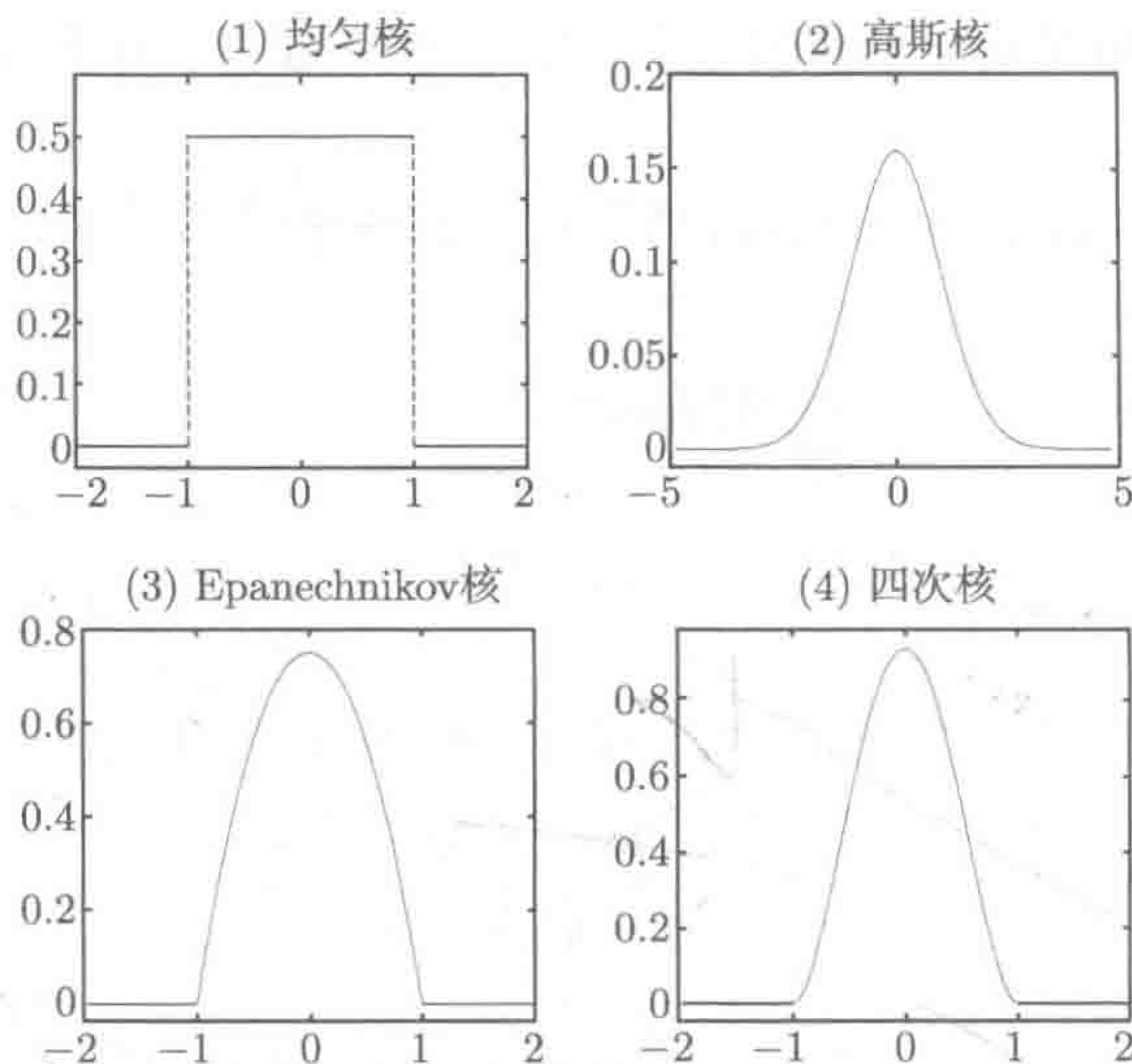


图 1.2.2 核函数曲线图

注 1.2.1 如果核函数 $K(\cdot)$ 取均匀核时, 式 (1.2.1) 所得的 N-W 核估计 $\hat{g}_{\mathrm{NW}}(x)$ 是落在邻域 $[x-h,x+h]$ 中的 X_i 对应的响应变量 Y_i 的算术平均值, 其中 $2h$ 表示这个邻域的宽度. 如果核函数 $K(\cdot)$ 取高斯核时, 式 (1.2.1) 所得的 N-W 核估计 $\hat{g}_{\mathrm{NW}}(x)$ 是 Y_i 的加权算术平均值, 当 X_i 落在离 x 越近时, 施加的权重就越大, 落在离 x 越远时, 施加的权重就越小, 当 X_i 落在 $[x-3h,x+3h]$ 之外时, 权重基本为 0.

2. *渐近偏差和渐近方差*

在一定条件下, 可以推得 N-W 核估计 $\hat{g}_{\mathrm{NW}}(x)$ 的渐近偏差和渐近方差分别为

$$\mathrm{bias}(\hat{g}_{\mathrm{NW}}(x))=\frac{1}{2}\left(g''(x)+\frac{2g'(x)f'(x)}{f(x)}\right)d_K h^2+o(h^2), \tag{1.2.2}$$

$$\mathrm{Var}(\hat{g}_{\mathrm{NW}}(x))=\frac{\sigma^2(x)c_K}{nhf(x)}+o((nh)^{-1}), \tag{1.2.3}$$

其中 $f(x)$ 表示 X 的密度函数, $d_K=\int_{-\infty}^{\infty}u^2K(u)\mathrm{d}u, c_K=\int_{-\infty}^{\infty}K^2(u)\mathrm{d}u.$

从式 (1.2.2) 定义的渐近偏差和式 (1.2.3) 定义的渐近方差可以看出, 光滑参数窗宽 h 对 N-W 核估计 $\hat{g}_{\mathrm{NW}}(x)$ 有显著的影响, 当窗宽 h 变小时, 参与加权平均的样本就变少, 这时偏差变小, 而方差则变大. 反之, 当窗宽 h 变大时, 参与加权平均的样本变多, 则偏差变大, 而方差变小. 所以在对非参数函数 $g(\cdot)$ 进行拟合时, 一个重要的问题就是选取合适的光滑参数, 即窗宽 h. 下面就窗宽 h 的选择进行一些讨论.

3. 窗宽的选择

由上述讨论, 选取窗宽 h 的一个自然的办法就是平衡偏差和方差, 即极小化均方误差 (mean squared error, MSE), 获得最优的窗宽 h. 由式 (1.2.2) 和式 (1.2.3), 简单计算可得

$$\begin{aligned}\text{MSE}(\hat{g}_{\text{NW}}(x)) &= E[\hat{g}_{\text{NW}}(x) - g(x)]^2 = [\text{bias}(\hat{g}_{\text{NW}}(x))]^2 + \text{Var}(\hat{g}_{\text{NW}}(x)) \\ &\approx C_V c_K (nh)^{-1} + \frac{1}{4} C_B^2 d_K^2 h^4,\end{aligned}$$

其中 C_V 和 C_B 是与核函数 $K(\cdot)$ 和窗宽 h 无关的量. 因此, 极小化 MSE, 可获得理论上的最优窗宽如下

$$h_{\text{opt}} = \left[\frac{C_V c_K}{C_B^2 d_K^2}\right]^{1/5} n^{-1/5} =: cn^{-1/5}. \tag{1.2.4}$$

如果取最优窗宽 h_{opt}, 则容易看到 N-W 核估计 $\hat{g}_{\text{NW}}(x)$ 在内点处的最优收敛速度为 $O(n^{-4/5})$. 另外, 从式 (1.2.4) 也可以看出, 最优窗宽 h_{opt} 依赖于 c 的大小, 而 c 与 $f(x), f'(x), g'(x), g''(x)$ 和 $\sigma^2(x)$ 有关, 这些量都是未知的, 因此需要估计这些未知量, 得到 c 的大小, 进而选取最优窗宽. 但在实际应用中, c 的估计不是件容易的事情, 因为对于非参数函数 $g(\cdot)$ 的一阶导数 $g'(\cdot)$ 和二阶导数 $g''(\cdot)$, 以及密度函数的导数 $f'(\cdot)$ 的估计有一定的难度, 涉及的问题比较多, 进一步它们估计量的收敛速度也非常慢, 严重导致 c 的估计具有较大的偏差.

下面给出在实际应用中广泛采用的两种窗宽选取方法: 交错核实 (cross validation, CV) 方法和广义交错核实 (generalized cross validation, GCV) 方法.

CV 方法是选择窗宽 h 的一种简单实用的方法, 通过极小化下面的 CV 目标函数获得窗宽 h 的选取

$$\text{CV}(h) = \frac{1}{n}\sum_{i=1}^{n}[Y_i - \hat{g}_h^{-i}(X_i)]^2 \omega(X_i), \tag{1.2.5}$$

其中 $\hat{g}_h^{-i}(X_i)$ 表示去掉第 i 个样本观测 (Y_i, X_i) 后得到的带有窗宽 h 的估计, $\omega(x)$ 为非负的权函数, 可取为 $\omega(x) = I(|x-0.5| \leqslant 0.4)$. 极小化式 (1.2.5) 定义的 CV 目标函数, 可得最优窗宽为 $\hat{h}_{\text{cv}} = \arg\min\limits_{h} \text{CV}(h)$.

CV 方法核心的思想是: 去掉样本中第 i 个样本点 (Y_i, X_i), 相当于去掉了 $\hat{g}_h^{-i}(X_i)$ 中使核权函数达到最大值的项 $W_{ni}(X_i)Y_i$, 从而排除了在观测点 $x = X_i$ 处过分夸大的作用, 有效提高了其他观测点的重要程度.

Craven 和 Wahha (1979) 进一步改进了 CV 方法计算量大等缺点 (因为 CV 方法需要拟合 n 条曲线), 提出了 GCV 的选取方法. 记

$$(\hat{g}_h(X_1), \cdots, \hat{g}_h(X_n))^{\text{T}} =: S_h Y,$$

其中 $\hat{g}_h(X_i)$ 是 $g(x)$ 的具有窗宽 h 的任意一个非参数函数的拟合曲线, S_h 表示仅依赖于变量 X 的 $n \times n$ 的帽子矩阵, $Y = (Y_1, \cdots, Y_n)^{\mathrm{T}}$, 则 GCV 的最优窗宽定义为

$$\hat{h}_{\text{gcv}} = \arg\min_h \mathrm{GCV}(h) = \arg\min_h \frac{\dfrac{1}{n}\displaystyle\sum_{i=1}^{n}[Y_i - \hat{g}_h(X_i)]^2}{[n^{-1}\mathrm{tr}(I_n - S_h)]^2}.$$

4. 核函数的选择

N-W 核估计 $\hat{g}_{\mathrm{NW}}(x)$ 除了依赖于窗宽 h 以外, 还依赖于核函数 $K(\cdot)$. Fan 和 Gijbels (1996) 指出在实际应用中, 对 N-W 核估计 $\hat{g}_{\mathrm{NW}}(x)$ 的精度起重要作用的仍然是窗宽 h, 核函数的作用不是太大.

把由式 (1.2.4) 定义的最优窗宽代入到 MSE 中, 计算得到

$$\mathrm{MSE}(K) = \frac{5}{4}(C_V^2 C_B c_K^2 d_K)^{2/5} n^{-4/5}.$$

极小化 MSE(K), 找最优的核函数 $K(\cdot)$. 从上式可看到, MSE(K) 关于核函数仅依赖于 $c_K^2 d_K$, 即

$$\left(\int_{-\infty}^{\infty} K^2(u)\mathrm{d}u\right)^2 \int_{-\infty}^{\infty} u^2 K(u)\mathrm{d}u.$$

因此, 极小化上式即可获得最优核函数为 Epanechnikov 核:

$$K(u) = \frac{3}{4}(1 - u^2) I(|u| \leqslant 1).$$

进一步, 关于 N-W 核估计的理论性质和详细讨论可见 Fan 和 Gijbels (1996), 薛留根 (2015). 关于核光滑方法的计算, 可使用 R 语言中的 "np" 程序包.

1.2.2 Gasser-Müller 光滑方法

对由式 (1.2.1) 定义的 N-W 核估计进行导数计算和研究它的渐近性质时, 估计量中的随机分母是有一定的困难. 为了解决这个问题, Gasser 和 Müller (1979) 提出了 Gasser-Müller 估计方法. 假设数据已经按照 X 变量进行了排序, 则 Gasser-Müller 估计定义为

$$\hat{g}_{\mathrm{GM}}(x) = \sum_{i=1}^{n} \widetilde{W}_{ni}(x) Y_i, \tag{1.2.6}$$

其中 $\widetilde{W}_{ni}(x)$ 为 Gasser-Müller 权函数, 定义为

$$\widetilde{W}_{ni}(x) = \frac{1}{h}\int_{s_{i-1}}^{s_i} K\left(\frac{x - s}{h}\right)\mathrm{d}s, \quad i = 1, \cdots, n,$$

其中 $s_0 = 0, s_n = 1, s_i = (X_{(i)} + X_{(i+1)})/2, 1 \leqslant i \leqslant n-1, X_{(i)}$ 为 X_i 的次序统计量, $1 \leqslant i \leqslant n$. $K(\cdot)$ 为 $[-1,1]$ 上有界且对称的概率密度函数, h 为收敛于 0 的窗宽, 窗宽 h 的选取可采用 1.2.1 节介绍的 CV 或 GCV 方法进行选择.

由式 (1.2.1) 和式 (1.2.6) 可以看出, N-W 核估计和 Gasser-Müller 估计都是在 x 的一个邻域 $[x-h, x+h]$ 内, 把非参数函数看成是一个局部常数进行逼近而得到的估计. 当 x 在一个适当的取值范围内变化时, 进而可以得到整个曲线的估计.

在一定条件下, 可以推得 Gasser-Müller 估计 $\hat{g}_{\mathrm{GM}}(x)$ 的渐近偏差和渐近方差分别为

$$\mathrm{bias}(\hat{g}_{\mathrm{NW}}(x)) = \frac{1}{2} g''(x) d_K h^2 + o(h^2), \tag{1.2.7}$$

$$\mathrm{Var}(\hat{g}_{\mathrm{NW}}(x)) = \frac{3}{2} \frac{\sigma^2(x) c_K}{nhf(x)} + o((nh)^{-1}), \tag{1.2.8}$$

其中 $f(x)$ 表示 X 的密度函数, $d_K = \int_{-\infty}^{\infty} u^2 K(u)\mathrm{d}u, c_K = \int_{-\infty}^{\infty} K^2(u)\mathrm{d}u$.

进一步, Gasser-Müller 估计 $\hat{g}_{\mathrm{GM}}(x)$ 的渐近 MSE 为

$$\mathrm{MSE}(\hat{g}_{\mathrm{GM}}(x)) = \left\{ \frac{1}{4} [g''(x)]^2 d_K^2 h^4 + \frac{3}{2} \frac{\sigma^2(x) c_K}{nhf(x)} \right\} [1 + o(1)].$$

1.2.3 局部多项式光滑方法

1. 局部多项式估计

Fan (1993) 与 Fan 和 Gijbels (1996) 提出的局部多项式光滑方法具有良好的性质: ① 可以减小 N-W 核估计的渐近偏差; ② 可以减小由 Gasser 和 Müller (1979) 提出的 Gasser-Müller 估计的渐近方差; ③ 对边界效应具有自适应性, 即在边界处的收敛速度和内点处的收敛速度相同, 具有非参数最优的收敛速度; ④ 具有良好的最小最大有效性; ⑤具有容易解释和计算的优点, 并适应于导数的估计. 本节就局部多项式光滑方法进行简单介绍.

假设回归函数 $g(\cdot)$ 在 x 的邻域内有连续的 d 阶导数, 由 Taylor 展式可近似回归函数 $g(\cdot)$ 如下

$$g(u) \approx \sum_{j=0}^{d} \frac{g^{(j)}(x)}{j!} (u-x)^j =: \sum_{j=0}^{d} \beta_j (u-x)^j, \tag{1.2.9}$$

其中 $\beta_j = g^{(j)}(x)/j!$, u 为 x 邻域内的点. 式 (1.2.9) 表示用一个多项式函数对回归函数 $g(\cdot)$ 进行局部逼近. 假设 $\{(Y_i, X_i), i = 1, \cdots, n\}$ 是来自模型 (1.1.4) 的 i.i.d.

的随机样本, 构造如下的加权最小二乘目标函数

$$\sum_{i=1}^{n}\left[Y_i-\sum_{j=0}^{d}\beta_j(X_i-x)^j\right]^2 K_h(X_i-x), \tag{1.2.10}$$

其中 $K_h(\cdot)=K(\cdot/h)/h, K(\cdot)$ 为核函数, h 为窗宽. 极小化加权最小二乘目标函数 (1.2.10), 即可得到 $\beta=(\beta_0,\beta_1,\cdots,\beta_d)^{\mathrm{T}}$ 的估计. 为了方便求解, 用向量和矩阵重新表示, 令 $\mathcal{W}=\operatorname{diag}\{K_h(X_1-x),\cdots,K_h(X_n-x)\}$ 为 $n\times n$ 的对角矩阵, $Y=(Y_1,\cdots,Y_n)^{\mathrm{T}}$, 且

$$\mathcal{X}=\begin{pmatrix}1 & X_1-x & \cdots & (X_1-x)^p\\ \vdots & \vdots & & \vdots\\ 1 & X_n-x & \cdots & (X_n-x)^p\end{pmatrix}.$$

加权最小二乘目标函数 (1.2.10) 可以重新写成下面矩阵形式

$$(Y-\mathcal{X}\beta)^{\mathrm{T}}\mathcal{W}(Y-\mathcal{X}\beta).$$

极小化上式, 可得 β 的加权最小二乘估计为

$$\hat{\beta}=(\mathcal{X}^{\mathrm{T}}\mathcal{W}\mathcal{X})^{-1}\mathcal{X}^{\mathrm{T}}\mathcal{W}Y. \tag{1.2.11}$$

由 $\beta_j=g^{(j)}(x)/j!$ 可知, 估计量 $\hat{\beta}$ 应该依赖于 x, 为了突出这个依赖关系, 进一步记为 $\hat{\beta}(x)=(\hat{\beta}_0(x),\hat{\beta}_1(x),\cdots,\hat{\beta}_d(x))^{\mathrm{T}}$, 则可得 $\hat{\beta}_j(x)=e_{j+1,d+1}^{\mathrm{T}}\hat{\beta}(x)$, 其中 $e_{j+1,d+1}$ 表示 $d+1$ 维的单位列向量, 第 $j+1$ 个元素为 1, 其余元素均为 0. 进一步, 可以获得 $g^{(j)}(x)$ 的估计为

$$\hat{g}^{(j)}(x)=j!\hat{\beta}_j(x), \quad j=0,1,\cdots,d. \tag{1.2.12}$$

需要注意的是, 回归函数 $g(x)$ 的局部多项式估计为

$$\hat{g}(x)=\hat{\beta}_0(x)=e_{1,d+1}^{\mathrm{T}}(\mathcal{X}^{\mathrm{T}}\mathcal{W}\mathcal{X})^{-1}\mathcal{X}^{\mathrm{T}}\mathcal{W}Y. \tag{1.2.13}$$

下面讨论一种特殊情况, 局部线性估计. 当 $d=1$ 时, 把由式 (1.2.13) 定义的非参数估计称为局部线性估计, 主要的原理是: 在 x 的一个邻域 $[x-h,x+h]$ 内, 用一个线性函数去逼近非参数函数 $g(x)$, 当 x 在适当的范围内变化时, 通过实施局部线性估计可以得到整个曲线 $g(\cdot)$ 的估计曲线, 直观的解释可见图 1.2.3. 经过计算, 可得非参数函数 $g(\cdot)$ 的局部线性估计为

$$\hat{g}_{\mathrm{LL}}(x)=\sum_{i=1}^{n}W_{ni}^{\mathrm{LL}}(x)Y_i, \tag{1.2.14}$$

其中

$$W_{ni}^{\mathrm{LL}}(x)=\frac{K_h(X_i-x)[S_{n,2}-(X_i-x)S_{n,1}]}{S_{n,0}S_{n,2}-S_{n,1}^2},$$

且$S_{n,k}=\sum_{i=1}^{n}(X_i-x)^kK_h(X_i-x), k=0,1,2$. 关于式 (1.2.14) 的窗宽选择, 可以采用 1.2.1 节介绍的 CV 方法或 GCV 方法进行选取. 关于局部多项式和局部线性估计的计算问题, 可以使用 R 语言中的 “KernSmooth” 软件包.

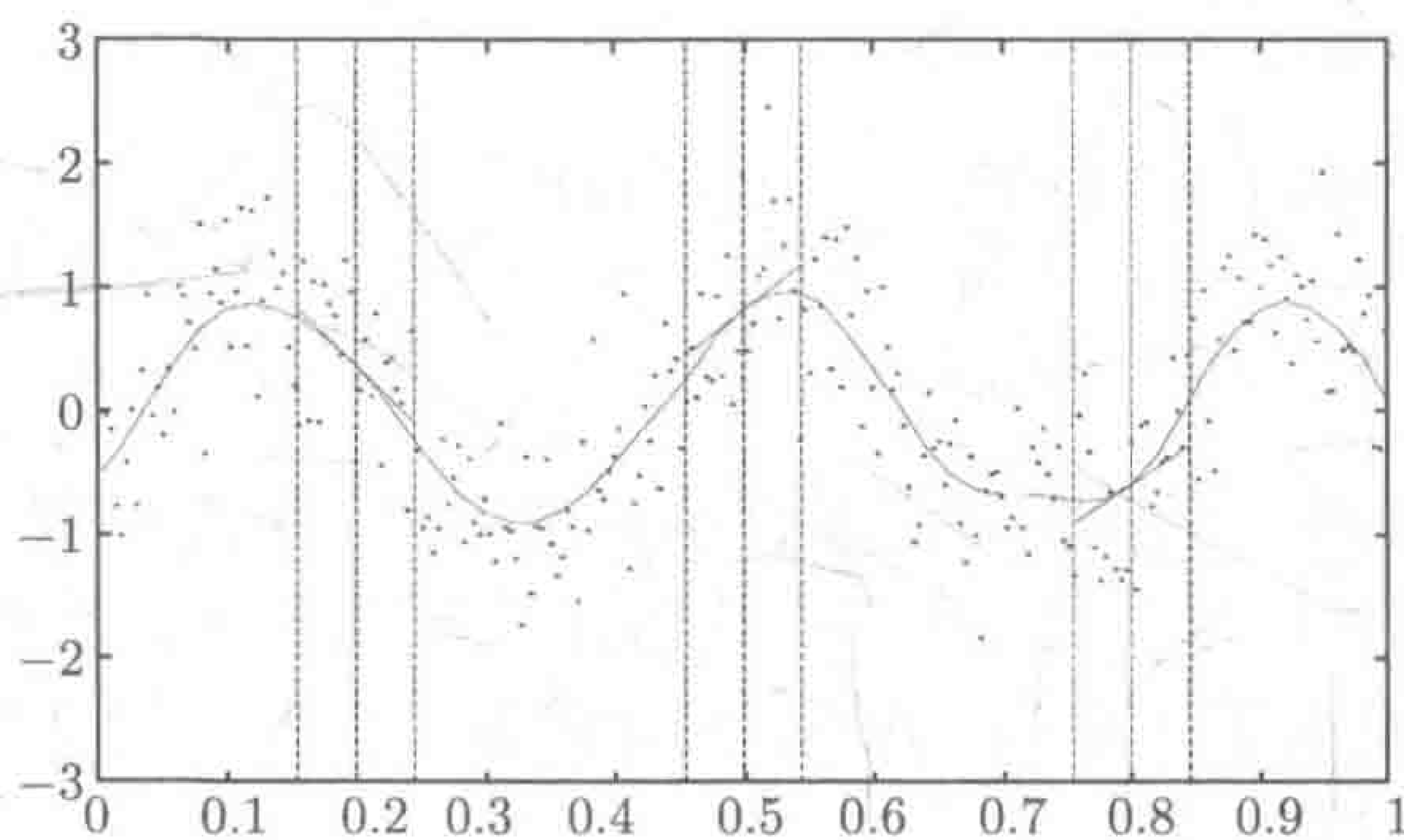


图 1.2.3 局部线性估计的原理示意图

2. 渐近条件偏差和方差

为了表示局部多项式估计的渐近偏差和渐近方差, 首先介绍一些概念. 令 $f(x)$ 表示变量 X 的概率密度函数, $\mathcal{F}_n$ 表示由 $\{X_i, i=1,\cdots,n\}$ 产生的 σ 代数. 进一步定义如下的矩阵和向量

$$S=(\mu_{j+l})_{0\leqslant j,l\leqslant d},\quad c_d=(\mu_{d+1},\cdots,\mu_{2d+1})^{\mathrm{T}},$$
$$\widetilde{S}=(\mu_{j+l+1})_{0\leqslant j,l\leqslant d},\quad \widetilde{c}_d=(\mu_{d+2},\cdots,\mu_{2d+2})^{\mathrm{T}},$$
$$S^*=(\nu_{j+l})_{0\leqslant j,l\leqslant d},$$

其中 $\mu_j=\int_{-\infty}^{\infty}u^jK(u)\mathrm{d}u, \nu_j=\int_{-\infty}^{\infty}u^jK^2(u)\mathrm{d}u, j=0,1,\cdots,d$.

当 $n\to\infty$ 时, $h\to 0, nh\to\infty$, 可以推得局部多项式估计 $\hat{g}^{(j)}(x)$ 的渐近条件方差为

$$\mathrm{Var}(\hat{g}^{(j)}(x)|\mathcal{F}_n)=e_{j+1,d+1}^{\mathrm{T}}S^{-1}S^*S^{-1}e_{j+1,d+1}\frac{j!^2\sigma^2(x)}{f(x)nh^{1+2j}}+o_P\left(\frac{1}{nh^{1+2j}}\right). \quad (1.2.15)$$

当 $d-j$ 为奇数时, 渐近条件偏差为

$$\mathrm{bias}(\hat{g}^{(j)}(x)|\mathcal{F}_n)=e_{j+1,d+1}^{\mathrm{T}}S^{-1}c_d\frac{j!g^{(d+1)}(x)}{(d+1)!}h^{d+1-j}+o_P(h^{d+1-j}). \quad (1.2.16)$$

进一步, 当 $d-j$ 为偶数时, 并且 $f'(x)$ 和 $g^{(d+2)}(x)$ 在 x 的邻域内连续, 如果 $nh^3 \to \infty$ 时, 则渐近条件偏差为

$$\begin{aligned}\text{bias}(\hat{g}^{(j)}(x)|\mathcal{F}_n) = & e_{j+1,d+1}^{\mathrm{T}} S^{-1} \widetilde{c}_d \frac{j!}{(d+2)!} \Big\{ g^{(d+2)}(x) \\ & + (d+2) g^{(d+1)}(x) \frac{f'(x)}{f(x)} \Big\} h^{d+2-j} + o_P(h^{d+2-j}). \quad (1.2.17)\end{aligned}$$

注 1.2.2　非参数函数 $g(x)$ 的局部线性估计 $\hat{g}_{\mathrm{LL}}(x)$ 的渐近条件偏差和方差分别为

$$\begin{aligned}\text{bias}(\hat{g}_{\mathrm{LL}}(x)|\mathcal{F}_n) &= \frac{1}{2}\mu_2 g''(x) h^2 + o_P(h^2), \\ \text{Var}(\hat{g}_{\mathrm{LL}}(x)|\mathcal{F}_n) &= \frac{\nu_0 \sigma^2(x)}{f(x)} \frac{1}{nh} + o_P((nh)^{-1}).\end{aligned}$$

可以发现, 局部线性估计的渐近条件偏差不依赖于密度函数, 表明局部线性估计是自适应的. 另外可以发现, 局部线性估计的渐近条件方差相同于 N-W 核估计的渐近方差, 但是小于 Gasser-Müller 估计的渐近方差. 进一步, 局部线性估计 $\hat{g}_{\mathrm{LL}}(x)$ 的渐近条件 MSE 为

$$\text{MSE}(\hat{g}_{\mathrm{LL}}(x)|\mathcal{F}_n) = \left\{ \frac{1}{4}\mu_2^2 [g''(x)]^2 h^4 + \frac{\nu_0 \sigma^2(x)}{f(x)} \frac{1}{nh} \right\} [1 + o_P(1)].$$

由上面渐近条件偏差和方差可知, 局部多项式估计有以下几个问题需要注意.

(1) 窗宽 h 的选择. 局部多项式估计对窗宽的选择很敏感, 如果窗宽 h 太大, 尽管可以减小条件方差, 但是产生了较大的偏差, 得到的估计较为光滑. 反之, 如果窗宽太小, 尽管可以减小条件偏差, 但是又会引起较大的方差, 导致估计较粗糙. 所以对于具体的模型, 选择合适的窗宽是至关重要的, 同样理论上也是通过平衡偏差和方差选取最优窗宽, 即极小化估计 $\hat{g}^{(j)}(x)$ 的均方误差 (MSE), 可得理论的最优局部窗宽为

$$h_{j,\mathrm{opt}} = C_{j,d}(K) \left\{ \frac{\sigma^2(x)}{[g^{(d+1)}(x)]^2 f(x)} \right\}^{1/(2d+3)} n^{-1/(2d+3)},$$

其中

$$C_{j,d}(K) = \left[\frac{(d+1)!^2 (2j+1) \displaystyle\int K_j^{*2}(u) \mathrm{d}u}{2(d+1-j) \left\{ \displaystyle\int u^{d+1} K_j^* \mathrm{d}u \right\}^2} \right]^{1/(2d+3)},$$

且 $K_j^*(u) = e_{j+1,d+1}^{\mathrm{T}} S^{-1} (1, u, \cdots, u^d)^{\mathrm{T}} K(u)$ 表示等价核. 对于某个非负权函数 $\omega(x)$, 通过极小化下面的加权积分均方误差 (mean integrated squared error, MISE)

$$\int_{-\infty}^{\infty} \Big\{ [\text{bias}(\hat{g}^{(j)}(x)|\mathcal{F}_n)]^2 + \text{Var}(\hat{g}^{(j)}(x)|\mathcal{F}_n) \Big\} \omega(x) \mathrm{d}x,$$

可得理论的全局最优窗宽为

$$h_{j,\text{opt}} = C_{j,d}(K)\left\{\frac{\int_{-\infty}^{\infty}\sigma^2(x)\omega(x)/f(x)\mathrm{d}x}{\int_{-\infty}^{\infty}[g^{(d+1)}(x)]^2\omega(x)\mathrm{d}x}\right\}^{1/(2d+3)} n^{-1/(2d+3)}.$$

上述最优窗宽在实际问题中不能直接被使用, 因为最优窗宽中包含未知的量, 如 $f(x), \sigma^2(x)$ 和 $g^{(d+1)}(x)$, 可利用 plug-in 方法把这些未知量的相合估计代入方可使用. 为了改进理论最优窗宽的使用, Fan 和 Gijbels (1996) 提出了两种窗宽选择方法: 辅助的残差平方选择方法和精炼的窗宽选择方法. 详细的讨论见 Fan 和 Gijbels (1996) 的第 4 章, 以及薛留根 (2015) 的第 4 章.

(2) 多项式阶数 d 的选择. 由于估计的渐近偏差和方差主要由窗宽来控制, 所以相对来说, 多项式阶数 d 的选择就没有窗宽的选择重要. 对于给定的窗宽 h, 较大的 d 虽然会减小偏差, 但是将导致方差以及计算量的增加. 相反, 取较小的 d 会增大偏差. 对于 $g^{(j)}(x)$ 的估计, Fan 和 Gijbels (1996) 推荐使用最低的奇数阶进行拟合, 即选择 $d = j + 1$.

(3) 核函数 $K(\cdot)$ 的选择. 核函数 $K(\cdot)$ 的选择相对来说没有那么重要, 用不同的核函数所得到的估计在数值上非常接近, 这个结论已经在理论上和计算上得到证实. 通常在均方误差准则下, 可获得最优核函数为 Epanechnikov 核: $K(u) = \frac{3}{4}(1-u^2)I(|u| \leqslant 1)$. 相关的理论结果见 Fan 和 Gijbels (1996) 中的定理 3.4.

1.3 经验似然

本书中多次把经验似然方法用于测量误差回归模型的研究, 为了让读者更好的阅读本书, 本节简单介绍由 Owen (1988, 1990) 提出的经验似然方法的思想和若干应用. 经验似然方法是 Owen (1988) 在完全样本下提出的一种非参数统计推断方法, 它有类似于 bootstrap 的抽样特性, 它通常对总体分布的类型要求比较弱, 仅仅知道一些附加信息 (如总体的一阶矩、二阶矩). 近年来, 许多学者已经将经验似然方法应用到各种统计模型及各种领域, 相关成果可参考 Owen (2001) 的专著*Empirical Likelihood*, 王启华 (2004) 关于经验似然统计推断的发展综述, 以及 Xue 和 Zhu (2010) 的专著*Empirical Likelihood in Nonparametric and Semiparametric Models.*

1.3.1 经验似然的思想

设 $X_1, \cdots, X_n$ 表示来自总体 $X \in \mathbb{R}^p$ 的一组i.i.d. 的随机样本, 且具有共同的

分布函数 F_0. 样本的经验累积分布函数为

$$F_n(A) = \frac{1}{n}\sum_{i=1}^{n} I(X_i \in A),$$

其中 $I(X_i \in A)$ 表示任一事件 $A \subseteq \mathbb{R}^p$ 的示性函数. 关于分布函数 F 的非参数似然函数可以定义为

$$L(F) = \prod_{i=1}^{n} F(\{X_i\}) = \prod_{i=1}^{n} \Big(F(X_i) - F(X_i-)\Big),$$

其中 $F(\{X_i\})$ 表示分布 F 在 X_i 处的概率质量. 当 $p=1$ 时, $F(\{X_i\}) = P(X = X_i), i = 1, \cdots, n$. Owen (2001) 中的定理 2.1 证明, 如果 $F \neq F_n$ 时, 有 $L(F) < L(F_n)$. 该结果表明, 经验累积分布函数 F_n 是 F 的极大似然估计.

类似于参数推断中参数似然比的定义, 当分布函数完全未知的情况下, 可以定义非参数似然比为

$$R(F) = \frac{L(F)}{L(F_n)}.$$

称 $R(F)$ 为分布函数 F 的经验似然比函数. 不同于参数似然比, 非参数似然比函数 $R(F)$ 中不包含未知参数, 一个自然的问题是如何使用该非参数似然比函数对参数进行统计推断. 注意到一些参数向量是分布函数 F 的泛函, 即 $\theta = T(F) \in \mathbb{R}^q$, 其中 $T(\cdot)$ 是分布函数 F 的某泛函, F 属于某分布类 $\mathcal{F}$. 为了对 $\theta = T(F)$ 作检验, Owen (1988) 定义了经验似然比检验统计量

$$R(\theta) = \sup_F \{R(F) | \theta = T(F), F \in \mathcal{F}\}.$$

经验似然比实际上是一种截面非参数似然比函数, 要求 F 在满足约束条件 $\theta = T(F)$ 下使非参数似然比达到最大, 该约束条件的引入, 即把非参数的经验似然比函数转换成了关于参数向量 θ 的极大截面似然比函数, 这样就可以用该经验似然比统计量对参数向量作假设检验, 区间估计和其他的各种统计推断, 这就是所谓的经验似然方法 (王启华, 2004).

为了计算经验似然比, 假设样本 X_i 之间没有结点, 进一步假设样本点 X_i 有概率质量 $p_i > 0, i = 1, \cdots, n$, 并记这样的分布函数为 F_p. 因此, 计算 $R(\theta)$ 就变成关于 $p_i, i = 1, \cdots, n$ 求极大值的问题, 即有

$$R(\theta) = \max\left\{\prod_{i=1}^{n}(np_i) \;\middle|\; \sum_{i=1}^{n} p_i = 1, p_i > 0, \theta = T(F_p)\right\},$$

其中 $L(F_n) = (1/n)^n$. 显然, Lagrange 乘子法可以解决这种带有约束条件的极值问题. Owen (1988, 1990) 证明经验似然比统计量渐近收敛到 χ^2_q 分布, 即所谓的 Wilks

定理, 其中 χ_q^2 表示自由度为 q 的 χ^2 分布. 利用该理论结果可以确定临界值 r_0, 进而可以作各种统计推断任务.

1.3.2 经验似然的若干应用

为了更好地理解经验似然方法, 本节介绍经验似然的几个应用.

1. 总体均值

Owen (1988, 1990) 提出经验似然方法后, 首先应用于总体均值的统计推断研究. 假设 $X_1, \cdots, X_n$ 表示来自总体 $X \in \mathbb{R}^p$ 的一组 i.i.d. 的随机样本, 并且总体 X 的分布函数为 F, 具有均值 $\mu = (\mu_1, \cdots, \mu_p)^{\mathrm{T}}$ 和满秩的协方差阵 Σ, 则 $\mu = T(F) = \int x\mathrm{d}F(x)$, 而 $T(F_p) = \int x\mathrm{d}F_p = \sum_{i=1}^n p_i X_i$. 因此, 可以定义总体均值 μ 的经验似然函数为

$$R(\mu) = \max\left\{\prod_{i=1}^n (np_i) \;\middle|\; \sum_{i=1}^n p_i = 1, p_i > 0, \sum_{i=1}^n p_i X_i = \mu\right\}. \tag{1.3.1}$$

如果考虑总体均值 μ 在 $X_1, \cdots, X_n$ 的凸包中时, 唯一的极大值存在. 由 Lagrange 乘子法, 可得

$$p_i = \frac{1}{n}\frac{1}{1 + \lambda^{\mathrm{T}}(X_i - \mu)}, \quad i = 1, \cdots, n, \tag{1.3.2}$$

其中 λ 为 Lagrange 乘子, 由下式确定

$$\sum_{i=1}^n \frac{X_i - \mu}{1 + \lambda^{\mathrm{T}}(X_i - \mu)} = 0.$$

Owen (1990) 证明 Lagrange 乘子 λ 的收敛速度为: $\|\lambda\| = O_P(n^{-1/2})$. 进一步, 由式 (1.3.1) 和式 (1.3.2), 可以定义总体均值 μ 的经验对数似然比函数, 并证明如下结论

$$\begin{aligned} l(\mu) &= -2\log(R(\mu)) = 2\sum_{i=1}^n \log\{1 + \lambda^{\mathrm{T}}(X_i - \mu)\} \\ &= n(\bar{X} - \mu)^{\mathrm{T}} S^{-1} (\bar{X} - \mu)^{\mathrm{T}} + O_P(n^{-1/2}), \end{aligned}$$

其中 $\bar{X} = \frac{1}{n}\sum_{i=1}^n X_i$ 和 $S = \frac{1}{n}\sum_{i=1}^n (X_i - \mu)(X_i - \mu)^{\mathrm{T}}$. 当 $n \to \infty$ 时, 显然有 $l(\mu) \xrightarrow{L} \chi_p^2$. 由此渐近结果, 可以构造总体均值 μ 的置信水平为 $1 - \alpha$ 的近似置信域如下

$$I_\alpha(\widetilde{\mu}) = \left\{\widetilde{\mu} \mid l(\widetilde{\mu}) \leqslant \chi_p^2(1 - \alpha)\right\},$$

其中 $\chi_p^2(1-\alpha)$ 为自由度为 p 的 χ^2 分布的 $1-\alpha$ 分位点, 满足 $P(\chi_p^2 \leqslant \chi_p^2(1-\alpha)) = 1-\alpha$.

2. 线性回归模型

Owen (1991) 首次将经验似然方法用于回归模型的研究, 进一步开创了经验似然方法的应用范围. 下面就经验似然方法在线性回归模型中的应用作简单回顾.

考虑下面的线性回归模型:

$$Y_i = X_i^{\mathrm{T}}\beta + \varepsilon_i, \quad i = 1, \cdots, n,$$

其中 $X_i = (X_{i1}, \cdots, X_{ip})^{\mathrm{T}}, \beta = (\beta_1, \cdots, \beta_p)^{\mathrm{T}}$ 为未知的回归参数向量, ε_i 是均值为 0 的模型误差. 对于线性回归模型, 由 $E(X(Y - X^{\mathrm{T}}\beta)) = 0$, 可以构造无偏的估计方程为

$$\sum_{i=1}^{n} X_i(Y_i - X_i^{\mathrm{T}}\beta) = 0.$$

通过求解上面的无偏的估计方程, 可以得到 β 的无偏的最小二乘估计. 因此可以定义辅助向量 $\eta_i(\beta) = X_i(Y_i - X_i^{\mathrm{T}}\beta)$, 构造 β 的经验似然函数为

$$R(\beta) = \max\left\{\prod_{i=1}^{n}(np_i) \;\middle|\; \sum_{i=1}^{n} p_i = 1, p_i \geqslant 0, \sum_{i=1}^{n} p_i\eta_i(\beta) = 0\right\}.$$

在一定条件下, 并当 $n \to \infty$ 时, Owen (1991) 证明 $-2\log(R(\beta)) \xrightarrow{L} \chi_p^2$. 该结果可用于构造回归参数 β 的置信水平为 $1-\alpha$ 的近似置信域如下

$$I_\alpha(\widetilde{\beta}) = \left\{\widetilde{\beta} \mid -2\log(R(\widetilde{\beta})) \leqslant \chi_p^2(1-\alpha)\right\},$$

其中 $\chi_p^2(1-\alpha)$ 为自由度为 p 的 χ^2 分布的 $1-\alpha$ 分位点, 满足 $P(\chi_p^2 \leqslant \chi_p^2(1-\alpha)) = 1-\alpha$.

3. 估计方程

设 $X_1, \cdots, X_n$ 是一个 p 维的 i.i.d. 的随机样本, 来自于未知分布函数 F, 存在一个 q 维参数向量 $\theta(F)$. 假设存在 $r \geqslant q$ 个独立无偏矩约束函数 $g_j(X, \theta)$, 使得 $E(g_j(X, \theta)) = 0, j = 1, 2, \cdots, r$. 写成向量形式为

$$g(X, \theta) = (g_1(X, \theta), \cdots, g_r(X, \theta))^{\mathrm{T}}, \quad E_F\{g(X, \theta)\} = 0, \tag{1.3.3}$$

其中 $X \in \mathbb{R}^p$, $\theta \in \mathbb{R}^q$, $g(\cdot, \cdot): \mathbb{R}^{p+q} \to \mathbb{R}^r$.

对半参数模型, 有关 θ 和 F 的信息可由式 (1.3.3) 所表示的独立无偏矩约束函数获得的情况下, Qin 和 Lawless (1994) 定义了经验对数似然比函数

$$l(\theta)=\max\left\{\log\prod_{i=1}^{n}np_i\;\middle|\;\sum_{i=1}^{n}p_i=1,\ p_i\geqslant 0,\ \sum_{i=1}^{n}p_ig(X_i,\theta)=0\right\}.$$

由 Lagrange 乘子法, 容易得到

$$l(\theta)=-\sum_{i=1}^{n}\log\left[1+\lambda^{\mathrm{T}}g(X_i,\theta)\right],$$

其中 Lagrange 乘子 $\lambda(\theta)=(\lambda_1(\theta),\cdots,\lambda_r(\theta))^{\mathrm{T}}$ 满足

$$\frac{1}{n}\sum_{i=1}^{n}\frac{g(X_i,\theta)}{1+\lambda^{\mathrm{T}}g(X_i,\theta)}=0.$$

实际上, $\log\prod_{i=1}^{n}np_i=\log\prod_{i=1}^{n}p_i-\log\prod_{i=1}^{n}\frac{1}{n}$ 可看成是 $(p_1,\cdots,p_n)^{\mathrm{T}}$ 到点 $\left(\frac{1}{n},\cdots,\frac{1}{n}\right)^{\mathrm{T}}$ 的似然距离. Qin 和 Lawless (1994) 证明了

$$-2l(\theta)=n\bar{g}^{\mathrm{T}}(\theta)\hat{\Sigma}^{-1}(\theta)\bar{g}(\theta)+o_P(1), \tag{1.3.4}$$

其中

$$\bar{g}(\theta)=\frac{1}{n}\sum_{i=1}^{n}g(X_i,\theta),\quad \hat{\Sigma}(\theta)=\frac{1}{n}\sum_{i=1}^{n}g(X_i,\theta)g^{\mathrm{T}}(X_i,\theta).$$

通过极小化 $-2l(\theta)$ 可以得到参数 θ 的估计 $\hat{\theta}$, 称为极大经验似然估计. 容易证明当 $r=q$ 时 (just-identified case), 有 $\hat{\theta}=\widetilde{\theta}$, 其中 $\widetilde{\theta}$ 是估计方程 $\frac{1}{n}\sum_{i=1}^{n}g(X_i,\theta)=0$ 的解, 并且在正则条件下, $\widetilde{\theta}$ 是唯一的 (Hansen, 1982). 当 $r>q$ 时 (over-identified case), $\frac{1}{n}\sum_{i=1}^{n}g(X_i,\theta)=0$ 无解. 这时, 由式 (1.3.4), 可利用由 Hansen (1982) 提出的广义矩方法 (generalized method of moments, GMM) 进行求解, 得到参数 θ 的估计. Qin 和 Lawless (1994) 在 $r>q$ 的情况下讨论了关于参数 θ, 分布 F 和 Lagrange 乘子 λ 的估计问题. 得到了 p_i 的估计, 即

$$\hat{p}_i=\frac{1}{n}\frac{1}{1+\hat{\lambda}^{\mathrm{T}}g(X_i,\hat{\theta})}.$$

分布 F 的估计为

$$\hat{F}_n(x)=\sum_{i=1}^{n}\hat{p}_iI(X_i<x).$$

Qin 和 Lawless (1994) 证明了即使 $r > q$ 时, 参数 θ 的极大经验似然估计 $\hat{\theta}$ 是存在的, 并且是唯一的. 他们还证明了参数 θ, 分布 F 和 Lagrange 乘子 λ 估计的渐近正态性, 并证明了经验对数似然比函数 $-2l(\theta)$ 的渐近 χ^2 分布. 李高荣和薛留根 (2007a, 2007b) 提出经验 Cressie-Read 统计量, 分别对独立同分布的数据和一般的弱相依数据进行了研究.

1.4 Bootstrap 方法

Efron (1979) 提出了一种经典的简单再抽样的统计方法 ——bootstrap 方法, 他在论文中系统地介绍了 bootstrap 方法和 bootstrap 分布. 在 bootstrap 方法提出之前, 统计应用中的再抽样方法有 Quenouille-Tukey Jackknife (刀切法). Jackknife 方法的基本思想是: 对于一个估计量, 不仅关心估计值的好坏, 也关心估计量的偏差和方差的大小, 即估计量的稳定性. 对于一个样本, 只能计算出一个估计值, 这时如何计算估计量的方差就成为一个主要问题. 同样, 参数真值未知时, 如何计算偏差也是一大难题. Jackknife 方法的做法是: 对于给定的一个样本, 每次删除一个 (或者几个) 样本点, 用剩下的样本点去重新计算估计值, 经过逐个删除并计算之后, 就可以得到一系列的估计值, 基于这一系列的估计值, 可以计算出偏差和方差.

基于 bootstrap 方法的理论基础, 使得 Efron (1979) 提出的 bootstrap 方法有了广泛的实际应用, 关于它的理论结果和应用的详细介绍见 Efron 和 Tibshirani (1993), 与 Hall (1992). Bootstrap 方法的核心是通过再抽样来构造 bootstrap 样本, 常用于数据的分布假设未知, 样本量很小, 需要解决的问题比较复杂, 构造的统计量不存在理论分布等情形.

下面对 bootstrap 方法做一个简单的描述. 假设 $X_1, \cdots, X_n$ 是 i.i.d. 的一组随机样本, 来自于一个未知的总体分布 F, 并令 $T_n = T_n(X_1, \cdots, X_n)$ 是感兴趣的一个统计量. 进一步, 假设 F_n 是样本 $X_1, \cdots, X_n$ 的经验分布函数, $X_1^*, \cdots, X_n^*$ 是从经验分布函数 F_n 抽取的一组随机样本. Bootstrap 方法就是在给定样本 $X_1, \cdots, X_n$ 的条件下, 通过 $T_n^* = T_n(X_1^*, \cdots, X_n^*)$ 的条件分布来估计 T_n 的分布. 条件分布称为 T_n 的 bootstrap 分布, T_n^* 称为 T_n 的 bootstrap 统计量. 具体过程如下:

步骤 1 由样本观测值构造样本的经验分布函数 F_n, 即 F_n 表示在每个样本点处具有概率质量 $1/n, i = 1, \cdots, n$. 事实上, 该经验分布正好是非参数的极大似然估计.

步骤 2 从经验分布函数 F_n 中抽取简单样本 $X_i^* = x_i^*, i = 1, \cdots, n$. 称 $X^* = (X_1^*, \cdots, X_n^*)$ 为 bootstrap 样本, 且 $X_i^* \sim F_n, i = 1, \cdots, n$. 这一步称为所谓的再抽样 (resample).

步骤 3 利用 bootstrap 样本 $X^* = (X_1^*, \cdots, X_n^*)$ 计算统计量 T_n, 记为 T_n^*.

步骤 4 对步骤 2~步骤 3 重复 B 次, 其中 B 要求为较大的正整数. 一般 B 的大小要根据实际检验问题和数据进行确定, 如果是通过统计量 T_n 构造置信区间, 一般 B 不能小于 1000.

步骤 5 利用 bootstrap 统计量 $T_{n,1}^*, \cdots, T_{n,B}^*$, 计算或画出该 bootstrap 统计量的频率直方图. 所得到的 bootstrap 分布就可以看成是 T_n 抽样分布的估计. 进而可以利用该分布对未知参数 θ 进行各种统计推断任务. 更多详细的讨论见 Efron 和 Tibshirani (1993) 与 Hall (1992).

1.5 变量选择方法

计算机的快速发展, 已经使得数据的收集变得越来越容易, 考虑的变量也越来越多, 问题也变得越来越复杂. 而在实际数据分析中, 仅有部分变量是重要的, 其他大部分变量都是不显著的. 因此, 为了提高模型的预测精度和模型的解释能力, 变量选择方法在统计建模和数据分析中变得越来越重要.

除了如 Akaike 信息准则 (AIC, Akaike, 1973) 和贝叶斯信息准则 (BIC, Schwarz, 1978) 等经典的 L_0 惩罚变量选择方法外, 一些收缩或 L_1 惩罚的变量选择方法成为统计学中重要的统计方法. 例如, Lasso (least absolute shrinkage and selection operator, Tibshirani, 1996), SCAD (smoothly clipped absolute deviation, Fan and Li, 2001) 和自适应的 Lasso (Zou, 2006). 尽管这些收缩的惩罚变量选择方法克服了经典 L_0 惩罚变量选择方法的计算量大和估计量的渐近性质研究困难等缺点, 但是这些方法也限制到了一些情形. 首先, 这些收缩的变量选择方法不能直接用于超高维数据的回归模型. 为了解决这类问题, 一些两步的变量选择方法被发展起来, 如, 第一步采用确定性独立筛选方法 (sure independence screening, SIS, Fan and Lv, 2008, Li et al., 2012b) 把超高维数 p 减小到小于样本量大小 n 的维数, 第二步使用收缩的惩罚变量选择方法同时进行变量选择和参数估计. 其次, 这些收缩的变量选择方法具有一些挑战性的问题, 如惩罚函数在原点是不可微的, 关于回归参数是非凹的, 且最终的参数估计关于截断参数的选择是非常敏感的. 因此, 不仅需要发展联合的算法找到最优解, 而且需要发展数据驱动的准则针对各种回归模型来选取截断参数.

收缩的惩罚变量选择方法遇到的一个主要问题是如何选择惩罚函数. Fan 和 Li (2001) 建议一个好的惩罚函数将导致具有三个性质的估计量:

(1) **无偏性** 当真参数很大时, 得到的估计量是渐近无偏的, 以避免不必要的建模偏差;

(2) **稀疏性** 所得到的估计量是一个门限值, 自动把小的参数分量估计成 0, 以便减少模型的复杂性;

(3) **连续性**　所得估计量在数据点处是连续的, 避免模型预测的不稳定性.

为了介绍惩罚变量选择方法, 介绍一些常用的惩罚函数. 首先考虑下面一般形式的惩罚最小二乘目标函数

$$\frac{1}{2}(z-\theta)^2+p_\lambda(|\theta|), \tag{1.5.1}$$

其中 $p_\lambda(\cdot)$ 为惩罚函数, λ 为截断参数或调整参数, 是用来控制模型的复杂度, 可以采用交错核实 (CV) 方法, 广义交错核实 (GCV) 方法或 BIC 等数据驱动的准则进行选取. 关于惩罚函数, Fan 和 Li (2001) 讨论了满足上面三个性质的惩罚函数所应满足的条件, 结论是: ①取值较大的真参数估计具有无偏性的充要条件是对取值较大的 $|\theta|$ 有 $p'_\lambda(|\theta|)=0$; ②具有稀疏性的充分条件是 $\min_\theta\{|\theta|+p'_\lambda(|\theta|)\}>0$; ③具有连续性的充要条件是 $\arg\min_\theta\{|\theta|+p'_\lambda(|\theta|)\}=0$. 可知, 满足稀疏性和连续性的惩罚函数一定在原点处奇异, 而非凸性的要求是减少估计的偏差. 因此, 在原点的奇异性是产生稀疏解的一个必要条件.

L_2 惩罚函数 $p_\lambda(|\theta|)=\lambda|\theta|^2$, 极小化惩罚最小二乘函数 (1.5.1), 得到的是岭回归 (ridge regression) 估计. 明显, L_2 惩罚函数在原点处不是奇异的, 因此 L_2 惩罚函数不能产生稀疏解. L_2 惩罚函数的一个推广形式是 L_q 惩罚函数 $p_\lambda(|\theta|)=\lambda|\theta|^q$, $q>1$. 这类惩罚函数只能减小估计的方差, 产生的是有偏估计, 但是不具有稀疏性.

L_1 惩罚函数 $p_\lambda(|\theta|)=\lambda|\theta|$ (图 1.5.1(b)), 极小化惩罚最小二乘函数 (1.5.1), 产生一个软门限解如下 (图 1.5.2(b))

$$\hat{\theta}=\operatorname{sgn}(z)(|z|-\lambda)_+. \tag{1.5.2}$$

Tibshirani (1996) 把 L_1 惩罚函数施加于回归模型的一般最小二乘和似然函数, 提出了 Lasso 变量选择方法. Lasso 变量选择方法尽管可以产生稀疏解, 并满足连续性, 但是所得估计不具有无偏性. 为了解决 Lasso 方法的缺陷, Zou (2006) 提出了自适应的 Lasso 方法, 保证了解的 oracle 性质. Knight 和 Fu (2000) 进一步把 L_1 惩罚函数推广到 L_q 惩罚函数 $q\leqslant 1$, 称为桥回归 (Frank and Friedman, 1993). 类似于 L_1 惩罚函数, L_q $(q\leqslant 1)$ 惩罚函数可以得到稀疏解, 但是同样不能保证大的真参数估计的无偏性.

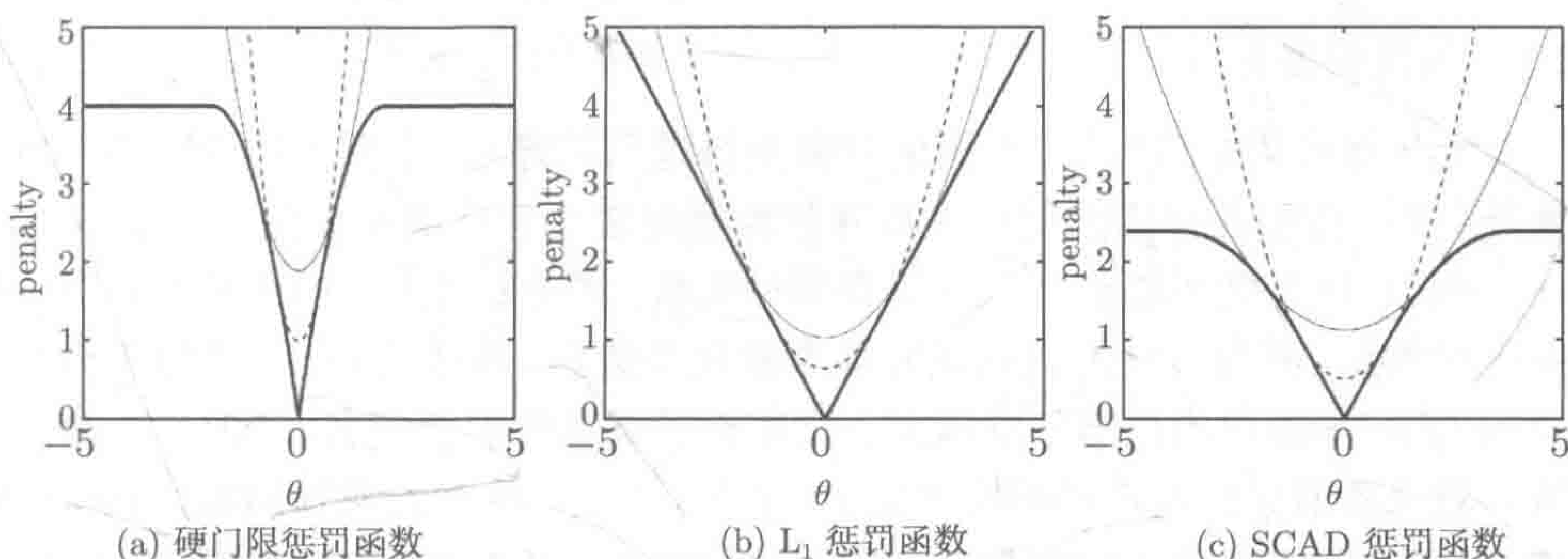
Antoniadis (1997) 提出了如下的硬门限惩罚函数 (图 1.5.1(a))

$$p_\lambda(|\theta|)=\lambda^2-(|z|-\lambda)^2I(|\theta|<\lambda). \tag{1.5.3}$$

如果施加硬门限惩罚函数, 可以得到如下的硬门限解 (Antoniadis, 1997; Fan, 1997)

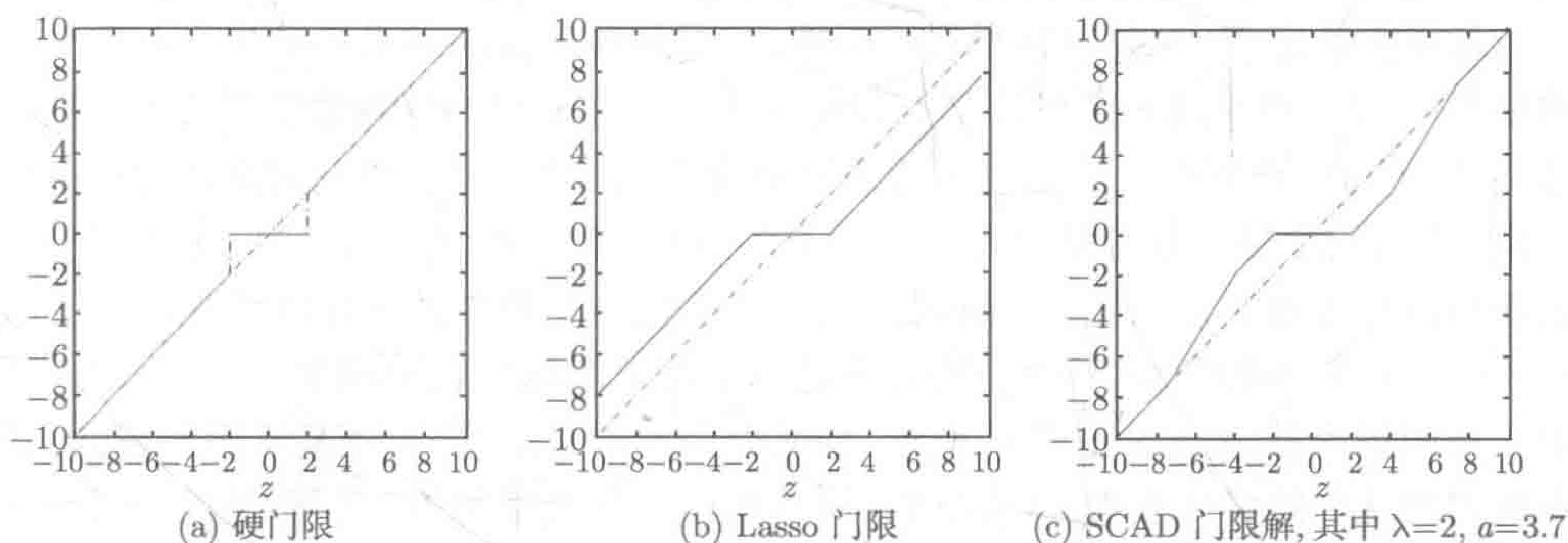
$$\hat{\theta}=zI(|z|>\lambda). \tag{1.5.4}$$

从图 1.5.2(a) 可以看出, 硬门限解满足无偏性和稀疏性, 但是对数据点 z, 不满足连续性.



(a) 硬门限惩罚函数 (b) L_1 惩罚函数 (c) SCAD 惩罚函数

图 1.5.1 三个惩罚函数 $p_\lambda(|\theta|)$ 和它们的二次逼近, $\lambda=2$, SCAD 惩罚函数中取 $a=3.7$



(a) 硬门限 (b) Lasso 门限 (c) SCAD 门限解, 其中 $\lambda=2$, $a=3.7$

图 1.5.2 门限函数图

上面提到的惩罚函数不能同时满足无偏性、稀疏性和连续性. 为了解决这个问题, Fan (1997) 提出了一个连续可微的惩罚函数, 称为 SCAD 惩罚函数 (图 1.5.1(c)), 定义为

$$p'_\lambda(|\theta|)=\lambda\left\{I(|\theta|\leqslant\lambda)+\frac{(a\lambda-|\theta|)_+}{(a-1)\lambda}I(|\theta|>\lambda)\right\}, \tag{1.5.5}$$

其中 $a>2$. Fan 和 Li (2001) 进一步对 SCAD 惩罚变量选择方法进行了研究, 证明 SCAD 惩罚函数可以同时满足无偏性、稀疏性和连续性, 并且具有 oracle 性质. Fan 和 Li (2001) 从 Bayes 角度建议取 $a=3.7$. 从图 1.5.1 可以看到, 不同于 L_1 惩罚函数, SCAD 惩罚函数有平坦的尾部, 这是减少偏差的基础, 即产生无偏估计的主要原因. 如果施加 SCAD 惩罚函数, 可以得到如下的 SCAD 惩罚最小二乘解 (图 1.5.2(c))

$$\hat{\theta}=\begin{cases}\operatorname{sgn}(z)(|z|-\lambda)_+, & |z|\leqslant 2\lambda,\\ \{(a-1)z-\operatorname{sgn}(z)a\lambda/(a-2)\}, & 2\lambda\leqslant|z|\leqslant a\lambda,\\ z, & |z|>a\lambda.\end{cases} \tag{1.5.6}$$

1.6　测量误差模型的估计方法

1.6.1　测量误差介绍

随着计算机的快速发展, 数据的收集变得越来越容易, 实际工作者用各种方式收集数据时, 由于各种原因使得收集到的数据经常会存在测量误差 (measurement error). 例如, 医生要判断患者是否患有某种疾病, 首先要对患者的生理指标进行检查并收集数据, 而收集到的数据经常是测量误差数据. 原因有很多, 有时因为涉及患者的隐私, 所提供的信息不准确或不完整而导致测量误差数据; 有时因为测量方法或测量仪器而导致测量误差的产生 (印赞华, 2013). 尽管收集到的数据存在测量误差, 但是实际工作者在对测量误差数据进行统计推断和分析时, 仍然希望能够正确分析变量之间的关系, 从而进行有效的统计推断和统计预测等.

长期实践证明, 如果忽略测量的影响, 将导致统计推断结果不准确或存在一定的偏差. 因此, 越来越多的统计学家和实际工作者认识到处理测量误差的重要性和必要性. 例如, 要判断一个患者是否患有高血压, 需要对患者在观测期内进行血压测量, 然后医生感兴趣的是观测期内的平均血压, 然而在实际问题中, 医生只能在某些时间点上进行血压测量, 而这些血压值还受到了测量时的仪器和测量血压的不同医生差异等的影响. 因此, 就会考虑一个简单的测量误差模型 $Y = y + \varepsilon$, 其中 ε 是测量误差, y 是所研究的变量, 如某个患者的血压值, 但不能准确测量到, 所观测到的是它的替代变量 Y. 把这种简单的测量误差模型推广到测量误差回归模型, 考虑响应变量和协变量之间的关系. 测量误差模型存在于许多领域, 像生物学、传染病学、森林、经济学和遥感等. 对于解释变量带有测量误差的回归问题一般称为带有变量误差模型或者 EV(errors-in-variables) 模型. 在测量误差模型方面已有不少文献讨论, 如 Fuller (1987) 对线性带有变量误差模型的研究做了概括和总结, Carroll 等 (1995) 对非线性 EV 模型的研究做了详细的讨论.

本节剩余部分以一元线性测量误差模型和非参数测量误差模型为例, 简单说明研究测量误差模型的重要性和必要性.

1.6.2　校正的估计方法

考虑如下的一元线性测量误差模型:

$$\begin{cases} Y_i = \beta_0 + \beta_1 X_i + \varepsilon_i, \\ W_i = X_i + U_i, \quad i = 1, \cdots, n, \end{cases} \tag{1.6.1}$$

其中 β_0 和 β_1 是未知参数, Y_i 是响应变量, X_i 是不能被准确测量的协变量, 也称为潜在变量 (latent variable), 而测量到的是受到误差 U_i 干扰后的观测变量 W_i, ε_i 是

模型误差, U_i 是测量误差. 通常把模型 (1.6.1) 称为一元线性测量误差模型或一元线性 EV (errors-in-variables) 模型. 为了方便看到测量误差对最小二乘估计的影响, 进一步假设 $(X_i, \varepsilon_i, U_i)^{\mathrm{T}}$ 是 i.i.d. 的, 并且相互独立, 来自于均值向量为 $(\mu_X, 0, 0)^{\mathrm{T}}$, 方差阵为 $\mathrm{diag}(\sigma_X^2, \sigma_\varepsilon^2, \sigma_{uu}^2)$ 的多元正态分布, 即

$$(X_i, \varepsilon_i, U_i)^{\mathrm{T}} \sim N\Big[(\mu_X, 0, 0)^{\mathrm{T}}, \mathrm{diag}(\sigma_X^2, \sigma_\varepsilon^2, \sigma_{uu}^2)\Big]. \tag{1.6.2}$$

由模型 (1.6.1) 和假设条件 (1.6.2), 可知 $(Y_i, W_i)^{\mathrm{T}}$ 来自一个二元正态分布, 其中均值为

$$E\{(Y, W)\} = (\mu_Y, \mu_W) = (\beta_0 + \beta_1\mu_X, \mu_X),$$

协方差阵为

$$\mathrm{Cov}(Y, W) = \begin{pmatrix} \mathrm{Var}(Y) & \mathrm{Cov}(Y, W) \\ \mathrm{Cov}(W, Y) & \mathrm{Var}(W) \end{pmatrix} = \begin{pmatrix} \beta_1^2\sigma_X^2 + \sigma_\varepsilon^2 & \beta_1\sigma_X^2 \\ \beta_1\sigma_X^2 & \sigma_X^2 + \sigma_{uu}^2 \end{pmatrix}.$$

用观测变量 W_i 代替潜在变量 X_i, 可得回归系数 β_1 的最小二乘估计为

$$\widetilde{\beta}_1 = \left[\sum_{i=1}^n (W_i - \bar{W})^2\right]^{-1} \sum_{i=1}^n (W_i - \bar{W})(Y_i - \bar{Y}), \tag{1.6.3}$$

其中 $\bar{W} = \frac{1}{n}\sum_{i=1}^n W_i$ 和 $\bar{Y} = \frac{1}{n}\sum_{i=1}^n Y_i$. 下面讨论该最小二乘估计是否是无偏的, 根据二元正态分布的假设, 可以计算得到

$$E(\widetilde{\beta}_1) = [\mathrm{Var}(W)]^{-1}\mathrm{Cov}(W, Y) = \beta_1(\sigma_X^2 + \sigma_{uu}^2)^{-1}\sigma_X^2. \tag{1.6.4}$$

明显可以看到, 当用观测变量 W_i 代替潜在变量 X_i 时, 得到的最小二乘估计 $\widetilde{\beta}_1$ 不是 β_1 的无偏估计, 并且受到了测量误差方差 σ_{uu}^2 的影响, 即受到了测量误差的影响而导致了估计的有偏性, 并且可以看到随着测量误差方差的变大, 而得到的最小二乘估计 $\widetilde{\beta}_1$ 会变小. 此外, 所得斜率 $\beta_1(\sigma_X^2 + \sigma_{uu}^2)^{-1}\sigma_X^2$ 和真实斜率 β_1 的正负性是一致的, 并且有不等式 $|\beta_1|(\sigma_X^2 + \sigma_{uu}^2)^{-1}\sigma_X^2 \leqslant |\beta_1|$ 可知, 偏差随着测量误差的方差 σ_{uu}^2 趋于 0 而衰减到 0. 为了消除测量误差对回归系数 β_1 的影响, 下面介绍两种校正方法.

(1) 引进可信度比 (reliability ratio) $\kappa_X = \sigma_W^{-2}\sigma_X^2 = (\sigma_X^2 + \sigma_{uu}^2)^{-1}\sigma_X^2$. 当可信度比 κ_X 已知时, 可以定义回归系数 β_1 的校正的无偏估计为

$$\hat{\beta}_1 = \kappa_X^{-1}\widetilde{\beta}_1.$$

在许多实际问题中, 可信度比 κ_X 的信息是已知的, 如在心理学、社会学和抽样调查等领域 (Fuller, 1987).

(2) 假设测量误差方差 σ_{uu}^2 已知时, 由 $\sigma_W^2 = \sigma_X^2 + \sigma_{uu}^2$ 和式 (1.6.3), 可定义校正的最小二乘估计为

$$\hat{\beta}_1 = (m_{WW} - \sigma_{uu}^2)^{-1} m_{WY},$$

其中 $m_{WW} = \dfrac{1}{n-1}\sum_{i=1}^{n}(W_i - \bar{W})^2$ 和 $m_{WY} = \dfrac{1}{n-1}\sum_{i=1}^{n}(W_i - \bar{W})(Y_i - \bar{Y})$ 分别表示样本方差和样本协方差. 如果没有假设 (Y_i, W_i) 来自于一个二元正态分布时, 即不涉及总体均值参数的估计时, 可用极大似然估计 $n^{-1}(n-1)m_{WW}$ 和 $n^{-1}(n-1)m_{WY}$ 分别替换 m_{WW} 和 m_{WY}, 获得校正的最小二乘估计.

有了斜率参数 β_1 的校正估计后, 进一步可以定义截距项 β_0 的估计为 $\hat{\beta}_0 = \bar{Y} - \hat{\beta}_1\bar{W}$. Fuller (1987) 讨论了估计量 $(\hat{\beta}_0, \hat{\beta}_1)^{\mathrm{T}}$ 的渐近正态性. 上面讨论是假定可信度比 κ_X 和测量误差方差 σ_{uu}^2 已知, 但在实际问题中它们往往是未知的. 可以发现, 如果可信度比 κ_X 未知时, 最主要的一个问题是估计测量误差方差 σ_{uu}^2, 因此在实际问题中可以通过重复测量或核实数据的方法获得 σ_{uu}^2 的相合估计.

本节的讨论主要是为了说明, 对于测量误差模型, 如果忽略测量误差将导致有偏估计, 会影响模型的预测和控制精度, 所以针对不同的测量误差类型和测量误差模型提出校正的估计方法是非常必要的.

1.6.3　SIMEX 方法

Cook 和 Stefanski (1994) 针对参数测量误差模型, 假设测量误差的分布为正态分布时, 首次提出了 SIMEX (Simulation-Extrapolation) 方法对模型中参数估计的偏差进行修正. SIMEX 方法能够形象演示测量误差对回归参数估计的影响, 主要分两步进行: 模拟步 (simulation step) 和外推步 (extrapolation step). 模拟步是用再抽样方法, 用原始数据加上模拟误差数据产生新数据, 然后对每组观测数据进行估计. 外推步是利用模拟步得到的估计, 找出估计的变化规律, 进而拟合出变化曲线, 再外推插值到 $\lambda = -1$ 时估计的值, 即得到所求的估计.

为了能更好的理解 SIMEX 方法, 用经典的一元线性测量误差模型 (1.6.1) 来演示这个方法的估计过程. 假设 $\{(Y_i, W_i), i = 1, \cdots, n\}$ 是 i.i.d. 的一组样本, 来自于模型 (1.6.1), 由 1.6.2 节的讨论已知, 当用 Y_i 和 W_i 进行回归时, 得到的回归系数 β_1 的估计渐近收敛到 $\beta_1(\sigma_X^2 + \sigma_{uu}^2)^{-1}\sigma_X^2$. SIMEX 方法的估计过程如下.

模拟步 (simulation step): 对于任意给定的 $\lambda > 0$, 由模拟再抽样重复增加均值为 0, 方差为 $\lambda\sigma_{uu}^2$ 的误差到 W_i, 即对于每一个 $i = 1, \cdots, n$, 产生一系列变量

$$W_{ib}(\lambda) = W_i + (\lambda\sigma_{uu}^2)^{1/2}U_{ib}, \quad b = 1, \cdots, B,$$

其中 $U_{ib} \sim N(0,1)$ 是一组与观测数据相互独立的 i.i.d. 的随机误差变量, B 是一个给定的任意大的正整数, $\lambda \in \Lambda = \{\lambda_1, \lambda_2, \cdots, \lambda_M\}$, 这里 Λ 表示 λ 取值的格子点,

一般 λ 的取值范围取为 [0,2]. 对所有的 $\lambda \in \Lambda = \{\lambda_1, \lambda_2, \cdots, \lambda_M\}$, 基于产生的新数据集 $\{(Y_i, W_{ib}(\lambda)), i = 1, \cdots, n, b = 1, \cdots, B\}$, 计算斜率 β_1 的最小二乘估计如下

$$\hat{\beta}_{1,b}(\lambda) = \left[\sum_{i=1}^{n}\left(W_{ib}(\lambda) - \bar{W}_b(\lambda)\right)^2\right]^{-1}\sum_{i=1}^{n}\left(W_{ib}(\lambda) - \bar{W}_b(\lambda)\right)(Y_i - \bar{Y}),$$

其中 $\bar{W}_b(\lambda) = \frac{1}{n}\sum_{i=1}^{n} W_{ib}(\lambda), b = 1, \cdots, B$. 进一步, 对所得估计 $\hat{\beta}_{1,1}(\lambda), \cdots, \hat{\beta}_{1,B}(\lambda)$ 作平均, 即

$$\hat{\beta}_1(\lambda) = \frac{1}{B}\sum_{b=1}^{B}\hat{\beta}_{1,b}(\lambda).$$

由假设条件, 进一步计算可知上式所定义的最小二乘估计是下面参数的相合估计

$$\mathcal{G}(\lambda) = \frac{\sigma_X^2}{\sigma_X^2 + (1+\lambda)\sigma_{uu}^2}\beta_1.$$

外推步 (extrapolation step): 该步的想法就是画出 $\mathcal{G}(\lambda)$ 关于 λ 的拟合曲线图, 找出估计的变化规律. 由 $\mathcal{G}(-1) = \beta_1$, 然后外推到 $\lambda = -1$, 产生外推 (SIMEX) 估计 $\hat{\beta}_1 = \hat{\beta}_1(-1)$, 详见图 1.6.1. 当外推函数取 $\mathcal{G}(\lambda) = \gamma_1 + \gamma_2(\gamma_3 + \lambda)^{-1}$ 时, Cook 和 Stefanski (1994) 证明所得的外推估计 $\hat{\beta}_1$ 是相合的. Carroll 等 (1996) 进一步研究了 SIMEX 估计的渐近性质.

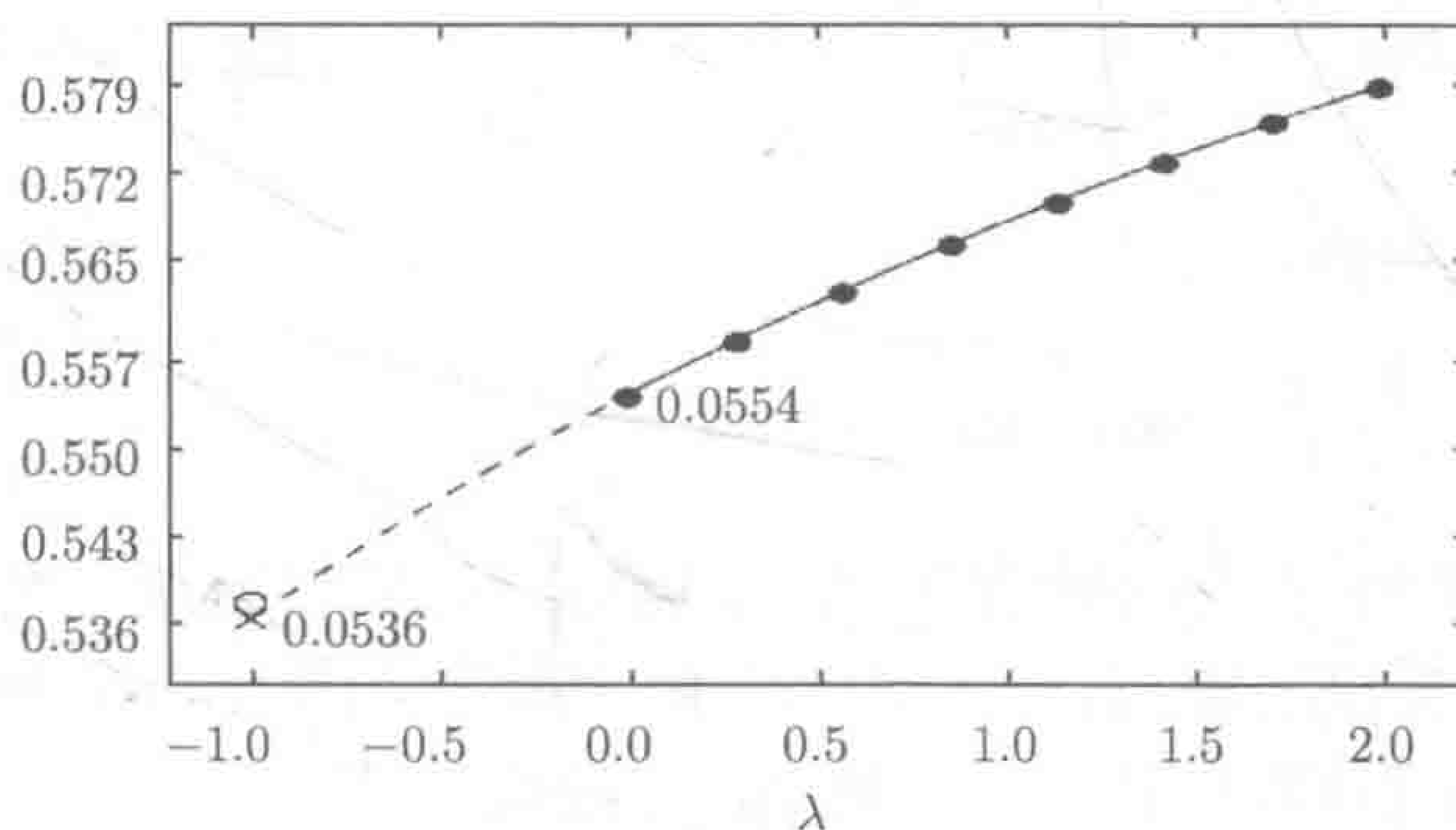


图 1.6.1 外推函数拟合示意图, $\lambda = 0$ 对应的是忽略测量误差得到的估计, $\lambda = -1$ 对应的是 SIMEX 估计

常用的外推函数有: ①线性外推函数, 即 $\mathcal{G}(\lambda) = \gamma_1 + \gamma_2\lambda$; ② 非线性外推函数, 即 $\mathcal{G}(\lambda) = \gamma_1 + \gamma_2(\gamma_3 + \lambda)^{-1}$; ③二次外推函数, 即 $\mathcal{G}(\lambda) = \gamma_1 + \gamma_2\lambda + \gamma_3\lambda^2$. Cook 和 Stefanski(1994) 证明, 当线性外推函数被使用时, SIMEX 估计的渐近偏差的阶为 $O(\sigma_{uu}^4)$; 而当非线性外推函数和二次外推函数被使用时, SIMEX 估计的渐近偏差的阶为 $O(\sigma_{uu}^6)$. 关于 SIMEX 方法的计算可使用 R 语言中的 simex 程序包.

1.6.4 逆卷积方法

本节考虑如下非参数测量误差模型:

$$\begin{cases} Y_i = g(X_i) + \varepsilon_i, \\ W_i = X_i + U_i, \end{cases} \tag{1.6.5}$$

其中 $g(\cdot)$ 是未知的光滑函数, $\{(Y_i, W_i), i = 1, \cdots, n\}$ 是 i.i.d. 的可观测的随机样本, X_i 是不可观测的协变量, ε_i 是模型误差, 满足 $E(\varepsilon_i|X_i) = 0$, U_i 是测量误差, 独立于 $(Y_i, X_i, \varepsilon_i)$, 并记 $f_U(u)$ 为误差变量 U_i 的密度函数. 当 X_i 带有测量误差时, 由 1.2.1 节介绍的核光滑方法不能直接用于 $g(\cdot)$ 的估计. 为了解决这个问题并构造非参数函数的相合估计, Fan 和 Truong (1993) 提出了逆卷积 (deconvolution) 的方法构造了 $g(\cdot)$ 的核估计, 并证明了核估计量可以达到局部和全局的最优收敛速度. 下面介绍 Fan 和 Truong (1993) 的估计方法.

令 $K(\cdot)$ 表示核函数, h 表示窗宽. 并令 $f_W(\cdot)$ 和 $f_X(\cdot)$ 分别表示变量 W 和 X 的概率密度函数. 由于 $W = X + U$, 可知 W 的密度函数 $f_W(\cdot)$ 是密度函数 $f_X(\cdot)$ 和 $f_U(\cdot)$ 的卷积, 即表示为

$$f_W(w) = \int_{-\infty}^{\infty} f_X(x) f_U(w - x) \mathrm{d}x.$$

令 $\phi_W(t)$ 表示 W 的特征函数, $\phi_X(t)$ 表示 X 的特征函数和 $\phi_U(t)$ 表示 U 的特征函数. 可以得到变量 X 的特征函数为

$$\phi_X(t) = \phi_W(t)/\phi_U(t).$$

由 Fourier 逆变换可得

$$f_X(x) = \frac{1}{2\pi} \int_{-\infty}^{\infty} \exp(-\mathrm{i}tx)\phi_X(t)\mathrm{d}t = \frac{1}{2\pi} \int_{-\infty}^{\infty} \exp(-\mathrm{i}tx)\frac{\phi_W(t)}{\phi_U(t)}\mathrm{d}t.$$

由假设条件可知, U 的密度函数 $f_U(\cdot)$ 已知, 而 W 的密度函数 $f_W(\cdot)$ 未知, 所以要想获得 X 的密度函数, 首先需要估计 W 的密度函数 $f_W(\cdot)$. 由于 $W_1, \cdots, W_n$ 是可观测变量, 因此可利用核方法得到 $f_W(\cdot)$ 的估计如下

$$\hat{f}_W(w) = \frac{1}{nh} \sum_{i=1}^{n} K\left(\frac{W_i - w}{h}\right).$$

进一步, 计算可得 $f_X(x)$ 的逆卷积核密度估计为

$$\begin{aligned} \hat{f}_X(x) &= \frac{1}{2\pi} \int_{-\infty}^{\infty} \exp(-\mathrm{i}tx)\frac{\hat{\phi}_W(t)}{\phi_U(t)}\mathrm{d}t \\ &= \frac{1}{2\pi} \int_{-\infty}^{\infty} \exp(-\mathrm{i}tx)\phi_K(th)\frac{\hat{\phi}_n(t)}{\phi_U(t)}\mathrm{d}t, \end{aligned} \tag{1.6.6}$$

其中 $\phi_K(\cdot)$ 表示核函数 $K(\cdot)$ 的 Fourier 变换, $\hat{\phi}_n(\cdot)$ 表示经验特征函数, 定义为

$$\hat{\phi}_n(t)=\frac{1}{n}\sum_{i=1}^{n}\exp(\mathrm{i}tW_i).$$

为了简单, 可以把式 (1.6.6) 定义的 $f_X(x)$ 的逆卷积核密度估计重新写成如下简单的核估计形式

$$\hat{f}_X(x)=\frac{1}{nh}\sum_{i=1}^{n}K_n\left(\frac{W_i-x}{h}\right), \tag{1.6.7}$$

其中

$$K_n(x)=\frac{1}{2\pi}\int_{-\infty}^{\infty}\exp(-\mathrm{i}tx)\frac{\phi_K(t)}{\phi_U(t/h)}\mathrm{d}t. \tag{1.6.8}$$

由式 (1.6.7) 和式 (1.6.8), Fan 和 Truong (1993) 定义了非参数函数 $g(\cdot)$ 的逆卷积核估计为

$$\begin{aligned}\hat{g}_n(x)&=\sum_{j=1}^{n}K_n\left(\frac{W_j-x}{h}\right)Y_j\Big/\sum_{i=1}^{n}K_n\left(\frac{W_i-x}{h}\right)\\&=\frac{1}{nh}\sum_{j=1}^{n}K_n\left(\frac{W_j-x}{h}\right)Y_j\Big/\hat{f}_X(x).\end{aligned} \tag{1.6.9}$$

通过下面的例子来演示逆卷积方法对非参数测量误差模型估计的有效性进行说明.

例 1.6.1 考虑下面的非参数测量误差模型 (1.6.5), 其中非参数函数取 $g(x)=4x(1-x)$, $\varepsilon_i\sim N(0,0.2^2)$, X_i 来自区间 $[0,1]$ 上的均匀分布, 测量误差 U_i 的密度函数为 $f_U(u)=(\sqrt{2}/3)^{-1}\exp(-3\sqrt{2}|u|)$. 图 1.6.2 给出了 Y 和 X 的散点图, 以及曲线拟合图. 图中实线表示真实曲线, 断线表示逆卷积方法的拟合曲线, 点断线表示忽略测量误差, 直接用观测样本 $\{(Y_i,W_i),i=1,\cdots,n\}$ 得到的拟合曲线.

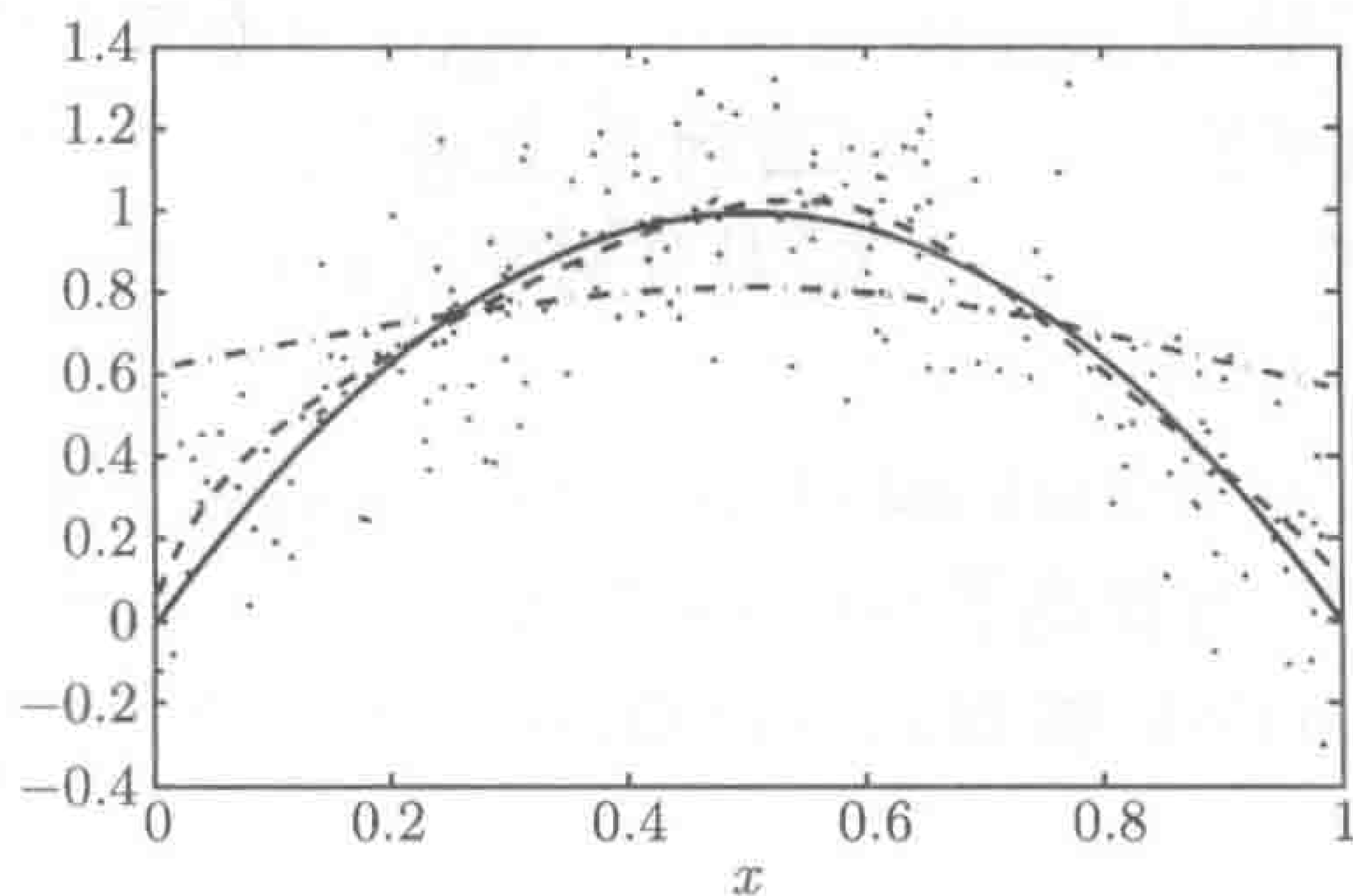


图 1.6.2 Y 和 X 的散点图, 以及曲线拟合图

从图 1.6.2 可以看出, 逆卷积方法的拟合效果基本接近真实曲线, 而忽略测量误差情形所得到的拟合曲线远离真实曲线, 可发现逆卷积方法有很好的效果.

Fan 和 Truong (1993) 详细讨论了由式 (1.6.9) 定义的估计量 $\hat{g}_n(x)$ 的相合性, 收敛速度等渐近性质, 感兴趣的读者可查阅该文献. Delaigle 和 Meister (2007) 将 Fan 和 Truong (1993) 的提出的逆卷积方法进行了推广, 推广到了异方差测量误差模型, 并讨论了估计量的渐近性质. 在上面讨论中, 假设测量误差变量 U 的密度函数 $f_U(u)$ 是已知的, 但在实际问题中通常是未知的, 可以通过重复测量的方法解决这个问题, 详细的讨论见 Delaigle 等 (2008). Delaigle 等 (2009) 进一步提出了基于重复测量的自适应局部多项式的估计方法, 并讨论了在不同的测量误差分布的情形下自适应局部多项式估计的渐近性质.

1.7　扭曲测量误差数据

在实际的医疗数据分析中, 响应变量和协变量通常会被某一共同的混淆变量的失真函数所污染, 观测到的是受到污染以后的响应变量和协变量, 把这类数据称为扭曲测量误差数据. Kaysen 等 (2003) 的纤维蛋白原数据就是一个该数据类型的典型例子, 其中血液透析患者的纤维蛋白原水平和血清铁传递蛋白水平均受身体质量指数 (body mass index, BMI) 的影响, BMI 定义为体重/身高^2. 另外一个例子就是对钙吸收量和钙摄取量之间关系的研究, Heaney 等 (1989) 和 Heaney (2003) 指出 BMI 对钙吸收量和钙摄取量均有影响. 这两个数据集的共同特点就是真实的响应变量和协变量都是不可观测的, 相应的, 人们所观测到的都是被一个共同的混淆变量污染过的数据.

下面介绍扭曲测量误差数据的特点. 假设 $X = (X_1, \cdots, X_p)^{\mathrm{T}}$ 表示 p 维的协变量向量, Y 表示真实的响应变量, 在实际问题中, 可观测的协变量和响应变量分别为

$$\widetilde{X}_r = \psi_r(U)X_r, \quad \widetilde{Y} = \phi(U)Y, \quad r = 1, \cdots, p,$$

其中 U 与 (Y, X) 相互独立, $\phi(\cdot)$ 和 $\psi_1(\cdot), \cdots, \psi_p(\cdot)$ 为混淆变量 U 的未知的一般光滑函数, 通常称为失真函数. 明显可以看到, $\widetilde{Y}$ 是被混淆变量 U 的未知失真函数 $\phi(\cdot)$ 污染后观测到的响应变量. $\widetilde{X}_1, \cdots, \widetilde{X}_p$ 是分别受到未知失真函数 $\psi_1(\cdot), \cdots, \psi_p(\cdot)$ 污染以后观测到的协变量.

受上述问题的启发, Sentürk 和 Müller (2005a) 对响应变量和协变量建立了协变量调整线性回归模型, 并提出了 binning 的估计方法. 考虑到模型的可识别性, 以及模型的估计程序, Sentürk 和 Müller (2005a) 对未知的光滑失真函数做了如下的约束

$$E\{\phi(U)\} = 1, \quad E\{\psi_r(U)\} = 1, \quad r = 1, \cdots, p.$$

自从 Sentürk 和 Müller (2005a) 的开创性工作以后, 许多统计学者对协变量调整回归模型进行了推广和发展, 更多的讨论见后面章节.

为了更好的理解扭曲测量误差数据的影响, 用一个简单的协变量调整回归模型进行说明.

例 1.7.1 考虑如下的模型:

$$
\begin{cases}
Y = 1 + 2X + \varepsilon, \\
\widetilde{Y} = \phi(U)Y, \\
\widetilde{X} = \psi(U)X,
\end{cases}
$$

其中 $\varepsilon \sim N(0, 0.5^2)$, $X \sim N(1,1)$, 混淆变量 U 来自区间 [1,7] 上的均匀分布, 失真函数取为 $\phi(U) = (-2U^2 + 4U + 90)/68$ 和 $\psi(U) = (U+1)/5$ 满足可识别性条件. 对于 $\widetilde{Y}$ 和 $\widetilde{X}$ 分别从相应的分布中抽取 100 个样本. 图 1.7.1(a) 给了真实变量 $\{(Y_i, X_i), i = 1, \cdots, 100\}$ 的散点图, 真实直线和基于真实变量得到的拟合直线. 图 1.7.1(b) 给了扭曲测量误差变量 $\{(\widetilde{Y}_i, \widetilde{X}_i), i = 1, \cdots, 100\}$ 的散点图, 真实直线和基于扭曲测量误差变量得到的拟合直线. 看到图 1.7.1(a) 中的拟合的直线非常接近于真实的直线, 然而图 1.7.1(b) 中的拟合的直线偏离真实直线很远. 扭曲测量误差变量的出现从本质上改变了 Y 和 X 的关系, 因而发展调整或校正的估计方法是非常必要的.

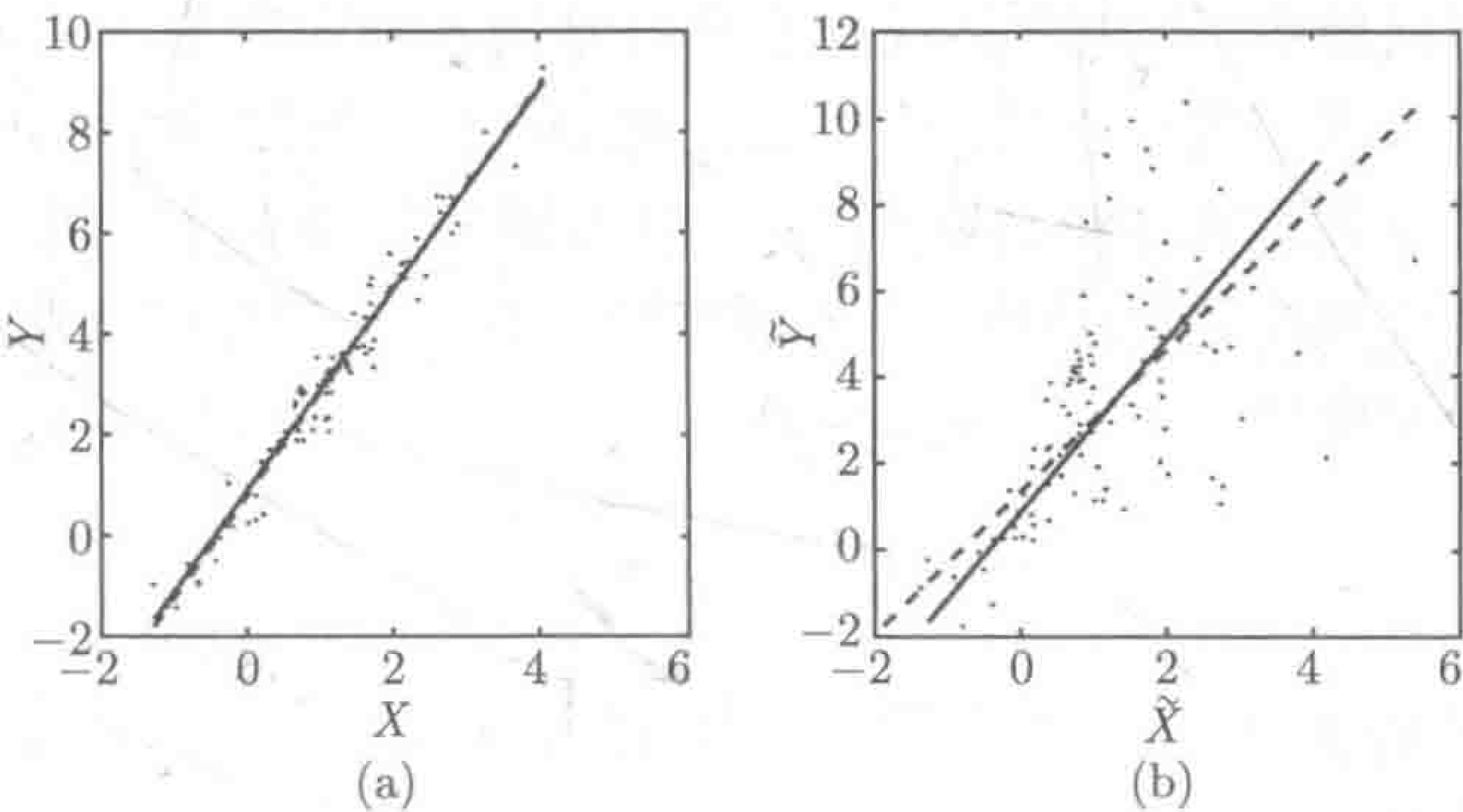


图 1.7.1 (a) Y 和 X 的散点图, 实线表示真实直线, 虚线表示拟合直线; (b) $\widetilde{Y}$ 和 $\widetilde{X}$ 的散点图, 实线表示真实直线, 虚线表示拟合直线

第 2 章　线性测量误差模型

本章主要针对线性可加测量误差模型和协变量调整线性回归模型, 讨论两种测量误差模型的估计方法和理论结果.

2.1　引　言

2.1.1　线性测量误差模型介绍

考虑经典的线性模型:

$$Y_i = X_i^{\mathrm{T}}\beta + \varepsilon_i, \quad i = 1, \cdots, n, \tag{2.1.1}$$

其中 $\beta = (\beta_1, \cdots, \beta_p)^{\mathrm{T}}$ 是 $p \times 1$ 的回归系数向量, X_i 是 p 维随机向量, ε_i 是均值为 0, 方差为 σ^2 的随机误差, $i = 1, \cdots, n$, 且 n 表示样本容量.

在很多实际问题中, 不能直接观测或准确测量到响应变量 Y_i 和协变量 X_i, 因此通常把带有变量误差的线性模型 (2.1.1), 称为线性 EV(errors-in-variables) 模型, 也称为线性测量误差模型. 对响应变量 Y 具有测量误差时比较简单, 可以把测量误差吸收到模型误差中进行处理. 首先考虑协变量 X_i 带有可加测量误差的情形, 即经典的测量误差模型:

$$W_i = X_i + U_i, \quad i = 1, \cdots, n, \tag{2.1.2}$$

其中 W_i 称为观测变量, X_i 称为潜在变量 (latent variable), U_i 是测量误差且与 (X_i, Y_i) 独立, 并且

$$U_1, U_2, \cdots, U_n \sim N(0, \Sigma_{uu}). \tag{2.1.3}$$

由模型 (2.1.1) 和模型 (2.1.2) 组成了线性 EV 模型, 详细讨论可见 Fuller (1987) 的专著. 一般考虑到模型 (2.1.1) 和模型 (2.1.2) 的可识别性, 要求测量误差的协方差阵 Σ_{uu} 已知或者要求 Σ_{uu} 与 σ^2 的比已知. 但在实际应用中这是不可能的, 这时可以通过重复测量的方法获得测量误差的协方差阵 Σ_{uu} 的相合估计. 关于 EV 回归模型可识别性的详细讨论见 Fuller (1987) 与 Carroll 等 (1995).

另外一种测量误差情况, 不能被直接观测的解释变量 X_i 和可观测变量 W_i 之间不是模型 (2.1.2) 所表示的可加关系, 而它们之间具有很复杂的关系. 即代替潜在变量 X_i, 所能观测到变量 W_i 满足如下关系

$$W_i = \phi(X_i, U_i), \quad i = 1, \cdots, n, \tag{2.1.4}$$

这里 U_i 是测量误差且与 (X_i, Y_i) 独立, $\phi(\cdot)$ 是一个任意已知函数. 在这种情况下要对未知参数 β 进行有效的统计推断是相当困难的. 利用核实数据可以克服这个困难, 不需要假定 $W_i = X_i + U_i$ 也能够使问题得到很好的解决. 在现实生活中, 诸如雇工的工资与工作时间的调查和学生的吸烟行为都是可获得核实数据的例子, 更多例子可见 Wittes 等 (1989), Pepe (1992), Sepanski 和 Lee (1995). 基于替代数据和核实数据的统计推断已引起国内外学者的重视. 例如: Sepanski 和 Lee (1995) 研究了基于核实数据的非线性 EV 模型; Wang 和 Rao (2002a) 利用经验似然方法构造了核实数据下线性 EV 模型中兴趣参数的置信域; 薛留根 (2006) 把经验似然方法推广到具有核实数据的非线性半参数 EV 模型, 构造了模型中未知参数的经验似然置信域; Wang (1999) 借助核实数据, 研究了部分线性 EV 模型的估计问题, 并证明了估计量的渐近性质; Wang (2003) 讨论了核实数据下部分线性 EV 模型的维数减少问题; 李高荣等 (2007, 2009) 分别在响应变量删失和缺失的情况下, 借助核实数据, 构造了线性 EV 模型中回归参数向量的经验似然置信域.

2.1.2 广义线性测量误差模型介绍

广义线性模型, 如伽马回归、逆高斯回归、泊松回归和 Logistic 回归, 在实际问题中有广泛的应用, 并吸引了许多统计学者和实际工作者的研究. 令 $\mu = E(Y|X)$, $\mathrm{Var}(Y|X) = V(\mu, \phi)$, 其中 ϕ 是一个已知的散度参数, $V(\cdot)$ 是一个已知的方差函数. 给定 X 条件下 Y 的广义线性模型为

$$g(\mu) = X^{\mathrm{T}}\beta, \tag{2.1.5}$$

其中 $g(\cdot)$ 是一个已知的联系函数, $\beta = (\beta_1, \cdots, \beta_p)^{\mathrm{T}}$ 是一个 $p \times 1$ 的未知回归参数向量. Wedderburn (1974) 研究了未知回归参数的拟似然估计, Firth (1987) 研究了拟似然估计的有效性问题. Chen 和 Cui (2003) 应用经验似然方法和额外的约束条件提高了回归参数估计的效率, 关于广义线性模型的详细讨论见 McCullagh 和 Nelder (1990), 李高荣和杨宜平 (2015) 的第 2 章.

然而在许多实际问题的研究中, 协变量不能够准确测量, 存在测量误差的情形. 如在艾滋病数据研究中 CD4 细胞数的测量 (Lin and Carroll, 2000; Liang, 2009), Framingham 心脏研究中血压的收缩压和高密度脂蛋白的测量都有测量误差 (Ma and Li, 2010; Yi, et al., 2011). 如果忽略测量误差, 所得的估计量和统计推断都是有

偏的. 令 W 是协变量 X 的具有测量误差的观测值. 假设如下的可加测量误差模型

$$W = X + U, \tag{2.1.6}$$

其中 $U \sim N(0, \Sigma_{uu})$ 独立于 (X, Y). 如果 X 的一些元素不具有测量误差, 即可以准确测量, 这时 U 相应的元素为零或者 Σ_{uu} 相应的方差分量为零. 测量误差问题已经被广泛研究, 如 Stefanski 和 Carroll (1987) 构造了广义线性测量误差模型的有效得分函数. Stefanski 和 Carroll (1989) 与 Nakamura (1990) 获得了诸如线性回归、伽马回归、逆伽马回归和泊松回归等广义线性模型的纠偏得分函数. Stefanski 和 Carroll (1989) 证明 Logistic 回归模型的纠偏得分函数不存在. Huang 和 Wang (2001) 研究 Logistic 回归模型中一些协变量存在测量误差, 提出了相合的泛函方法, 证明了估计量的渐近性质. Yang 等 (2015) 基于纠偏得分函数的矩识别条件, 提出了纠偏的经验似然方法, 研究了高斯回归、泊松回归和 Logistic 回归等广义线性测量误差模型, 在一定的正则条件下, 证明了经验对数似然比渐近收敛到标准 χ^2 分布, 所得结果可以构造回归系数的置信域.

Liang 等 (1999), Li 和 Xue (2008) 与 Liang 和 Li (2009) 研究了部分线性测量误差模型. Zhou 和 Liang (2009) 与 Zhang 等 (2011) 研究了半参数变系数部分线性测量误差模型的半参数 profile 最小二乘估计问题. Yi 等 (2011), 与 Yi 等 (2012) 讨论了协变量具有测量误差, 响应变量缺失时, 纵向数据模型的偏差分析和统计推断等问题. 目前, 测量误差模型的变量选择问题也成为统计学者研究的问题, 如 Liang 和 Li (2009), Ma 和 Li (2010), Zhao 和 Xue (2010).

2.1.3　协变量调整线性回归模型介绍

协变量调整回归模型最初是由 Sentürk 和 Müller (2005a) 在分析健康科学数据时, 受到启发而提出并发展起来的一类统计模型, 这类模型的研究已经成为统计学研究的热点问题之一. 首先介绍下面的协变量调整线性回归模型:

$$\begin{cases} Y = \beta_0 + \sum_{r=1}^{p} X_r\beta_r + \varepsilon, \\ \widetilde{Y} = \phi(U)Y, \\ \widetilde{X}_r = \psi_r(U)X_r, \quad r = 1, \cdots, p, \end{cases} \tag{2.1.7}$$

其中 ε 是独立同分布的随机误差, 且有 $E(\varepsilon) = 0$ 和 $\mathrm{Var}(\varepsilon) = \sigma^2$. Y 是不可观测的响应变量, 而观测到的是被混淆变量 U 的未知失真函数 $\phi(\cdot)$ 污染后的变量 $\widetilde{Y}$. $X = (X_1, \cdots, X_p)^{\mathrm{T}}$ 是不可观测的连续型协变量向量, $\widetilde{X}_r$ 是受到未知失真函数 $\psi_r(\cdot)$ $(r = 1, \cdots, p)$ 污染以后观测到的协变量, 用 $\widetilde{X} = (\widetilde{X}_1, \cdots, \widetilde{X}_p)^{\mathrm{T}}$ 表示扭曲污染后的可观测协变量向量. 未知的失真函数 $\phi(\cdot)$ 和 $\psi_1(\cdot), \cdots, \psi_p(\cdot)$ 允许假定为变量 U 的一般光滑函数.

考虑到模型的可识别性, 以及模型的估计程序, 需要对未知的光滑失真函数做如下的约束

$$E\{\phi(U)\} = 1, \quad E\{\psi_r(U)\} = 1, \quad r = 1, \cdots, p. \tag{2.1.8}$$

式 (2.1.8) 暗示着失真效应的均值为 0 (Sentürk and Müller, 2005a). 对于 $r = 1, \cdots, p$, 进一步假设 (X_r, U, ε) 是相互独立的, 且 (Y, U) 也是独立的.

自从协变量调整回归模型被提出以来, 近几年已经引起了统计学者的广泛关注. 例如, Sentürk 和 Müller (2005a) 把模型 (2.1.7) 转换成一个变系数模型, 基于变系数模型和装箱 (binning) 的思想, 利用局部光滑方法得到系数函数的估计, 然后利用失真函数的约束条件 (2.1.8), 进而得到截距参数 β_0 和未知参数向量 $\beta = (\beta_1, \cdots, \beta_p)^{\mathrm{T}}$ 的估计. 进一步, Sentürk 和 Müller (2006) 证明了估计量的渐近性质. Sentürk 和 Müller (2005b) 对模型 (2.1.7) 的相关性进行了分析, 并证明了相关的理论结果. 赵培信和薛留根 (2011) 提出了校正的经验似然方法, 构造了模型中回归参数的经验似然置信域. Binning 技巧被一些统计学者推广到了协变量调整变系数模型和广义协变量调整模型的研究, 见 Sentürk (2006), Sentürk 和 Müller (2006), Sentürk 和 Nguyen (2006), Sentürk 和 Müller (2009).

Cui (2008) 提出了一个自然的直接估计程序对模型 (2.1.7) 进行研究: 首先对响应变量和预测变量关于混淆变量 U 做回归, 采用非参数估计方法分别获得失真函数 $\phi(\cdot)$ 和 $\psi_r(\cdot)(r = 1, \cdots, p)$ 的相合估计, 然后获得不可观测响应变量和预测变量的估计记为 $(\hat{Y}, \hat{X}_1, \cdots, \hat{X}_p)$. Li 等 (2014b) 考虑了协变量的维数 p 随着样本量 n 的变大而发散的模型 (2.1.7), 提出了 SCAD 惩罚的最小二乘变量选择方法, 所提方法能够同时进行变量选择和参数估计, 并证明了估计量的渐近性质. Cui (2009), Zhang 等 (2012a) 推广这种自然估计的方法到协变量调整的非线性模型, 并从理论上进行了系统的研究. Zhang 等 (2013b) 进一步研究了具有失真测量误差的部分线性单指标模型.

2.1.4 本章结构

本章的内容主要有以下几个方面. 2.2 节介绍线性可加测量误差模型的三种估计方法: 校正的估计方法、M 估计方法和 SIMEX (Simulation-Extrapolation) 估计方法. 2.3 节针对线性测量误差模型 (2.1.1) 和 (2.1.4), 响应变量随机缺失的情况下, 借助于核实数据, 研究了回归系数和响应均值的经验似然推断问题. 2.4 节针对广义线性测量误差模型 (2.1.5) 和 (2.1.6), 基于纠偏的得分函数的矩可识别性, 提出纠偏的经验似然方法, 并证明回归参数的纠偏经验对数似然比统计量渐近收敛到标准 χ^2 分布, 同时证明回归参数极大经验似然估计的渐近正态性. 2.5 节针对协变量调整线性回归模型 (2.1.7), 介绍几种流行的估计方法, 如 binning 方法, 最近邻

binning 方法, 局部线性估计方法和直接估计方法. 2.6 节介绍回归参数置信域的构造方法, 如渐近正态近似方法、bootstrap 重抽样近似方法和经验似然方法. 2.7 节研究具有发散参数的协变量调整线性回归模型的变量选择问题, 提出了 SCAD 惩罚最小二乘方法同时进行变量选择和参数估计. 2.8 节主要研究协变量调整回归模型的拟合优度检验问题, 提出一种经验过程的检验统计量对回归函数的线性结构进行拟合优度检验.

2.2　线性测量误差模型的估计方法

本节针对线性 EV 模型 (2.1.1) 和 (2.1.2), 介绍三种估计方法: 校正的估计方法、M 估计方法和 SIMEX 估计方法.

2.2.1　校正的估计方法

如果在模型 (2.1.1) 中协变量 X 是可准确观测时, 则真实的回归参数向量 β 可以定义为下面总体的最小二乘估计

$$\beta = \arg\min_{\beta} E(Y - X^{\mathrm{T}}\beta)^2 = [E(XX^{\mathrm{T}})]^{-1}E(XY). \tag{2.2.1}$$

但是随机变量 X 不能准确测量, 可测量的是具有测量误差 U 的变量 W, 如果直接用 W 代替式 (2.2.1) 中的 X. 由测量误差 U 和 (Y, X) 相互独立性, 以及假设 (2.1.3), 则有

$$\beta^* = \arg\min_{\beta} E(Y - W^{\mathrm{T}}\beta)^2 = [E(XX^{\mathrm{T}}) + \Sigma_{uu}]^{-1}E(XY). \tag{2.2.2}$$

明显可以看到, 当用具有测量误差变量 W 代替 X 时, 得到的最小二乘估计 β^* 不是真实的回归参数向量 β. 为了消除测量误差的影响, 可以定义校正的最小二乘目标函数: $E(Y - W^{\mathrm{T}}\beta)^2 - \beta^{\mathrm{T}}\Sigma_{uu}\beta$, 其中减去一项 $\beta^{\mathrm{T}}\Sigma_{uu}\beta$, 主要目的是衰减由测量误差带来的影响, 保证校正的最小二乘目标函数在真实参数处衰减到 0, 使得求解得到的最小二乘估计等于真实的回归参数向量 β.

假设$\{(Y_i, W_i), i = 1, \cdots, n\}$ 是来自于模型 (2.1.1) 和 (2.1.2) 的 i.i.d. 的随机样本, 并且假设 Σ_{uu} 已知. 由上面的讨论, 则可得到回归参数向量 β 校正的最小二乘估计如下

$$\begin{aligned}\hat{\beta} &= \arg\min_{\beta}\left\{\frac{1}{n}\sum_{i=1}^{n}(Y_i - W_i^{\mathrm{T}}\beta)^2 - \beta^{\mathrm{T}}\Sigma_{uu}\beta\right\} \\ &= \left(\frac{1}{n}\sum_{i=1}^{n}W_iW_i^{\mathrm{T}} - \Sigma_{uu}\right)^{-1}\frac{1}{n}\sum_{i=1}^{n}W_iY_i.\end{aligned}$$

进一步, 因为 $E(Y_i-X_i^{\mathrm{T}}\beta)^2=\sigma^2$ 和 $E(Y_i-W_i^{\mathrm{T}}\beta)^2=E(Y_i-X_i^{\mathrm{T}}\beta)^2+\beta^{\mathrm{T}}\Sigma_{uu}\beta$, 并由校正的最小二乘估计 $\hat{\beta}$, 可得到模型误差的方差 σ^2 的估计如下:

$$\hat{\sigma}^2=\frac{1}{n}\sum_{i=1}^{n}(Y_i-W_i^{\mathrm{T}}\hat{\beta})^2-\hat{\beta}^{\mathrm{T}}\Sigma_{uu}\hat{\beta}.$$

关于估计量 $\hat{\beta}$ 和 $\hat{\sigma}^2$ 的大样本性质, 如相合性和渐近正态性, 在一些条件下是很容易证明的, 此处不再讨论.

在许多实际问题中, Σ_{uu} 已知的假定是不成立的, 因此需要估计 Σ_{uu} 的大小. Liang 等 (1999) 建议采用重复测量的方法来估计测量误差的协方差阵 Σ_{uu}. 假定重复测量, 即 $W_{ij}=X_i+U_{ij}, j=1,\cdots,m_i$. 令 $\bar{W}_i=\dfrac{1}{m_i}\sum_{j=1}^{m_i}W_{ij}$ 表示重复测量的样本均值, 则 Σ_{uu} 的一个无偏的矩估计可以定义为

$$\hat{\Sigma}_{uu}=\frac{\sum_{i=1}^{n}\sum_{j=1}^{m_i}(W_{ij}-\bar{W}_i)(W_{ij}-\bar{W}_i)^{\mathrm{T}}}{\sum_{i=1}^{n}(m_i-1)}. \tag{2.2.3}$$

2.2.2 M 估计方法

针对线性 EV 模型 (2.1.1) 和 (2.1.2), 崔恒建 (1995a) 假设 $(\varepsilon,U^{\mathrm{T}})^{\mathrm{T}}$ 为 $p+1$ 维球对称误差向量, 即 $(\varepsilon,U^{\mathrm{T}})^{\mathrm{T}}\overset{L}{=}RE_{p+1}$, 其中 $\overset{L}{=}$ 表示同分布, R 为非负随机向量, E_{p+1} 为 $\Omega_p=\{a|a\in\mathbb{R}^{p+1},\|a\|=1\}$ 上的均匀随机向量, 并且 R 与 E_{p+1} 独立. 崔恒建 (1995a) 在误差向量服从 $p+1$ 维球对称分布时, 提出了稳健的 M 估计方法, 并证明了回归参数向量 M 估计的强相合性和渐近正态性, 同时得到了 σ^2 估计的渐近性质. 崔恒建 (1997) 进一步提出广义最小一乘估计方法, 同样证明了最小一乘估计的相合性和渐近正态性.

假设$\{(Y_i,W_i),i=1,\cdots,n\}$ 是来自于模型 (2.1.1) 和 (2.1.2) 的 i.i.d. 的随机样本, 选取适当的 $\rho(\cdot)$ 函数, 定义如下的 M 估计

$$\hat{\beta}_n=\arg\min_{\beta}\left\{\frac{1}{n}\sum_{i=1}^{n}\rho\left(\frac{Y_i-W_i^{\mathrm{T}}\beta}{\sqrt{1+\|\beta\|^2}}\right)\middle|\beta\in\mathbb{R}^p\right\}. \tag{2.2.4}$$

进一步, 由式 (2.2.4) 所定义的 M 估计 $\hat{\beta}_n$, 可以定义 σ^2 的估计如下

$$\hat{\sigma}_n^2=\frac{1}{n}\sum_{i=1}^{n}\frac{(Y_i-W_i^{\mathrm{T}}\hat{\beta}_n)}{1+\|\hat{\beta}_n\|^2}.$$

Cui 和 Chen (2003) 提出了经验似然方法, 构造线性 EV 模型中回归系数 β 的置信域, 并证明了经验对数似然比统计量的非参数 Wilks 定理, 他们进一步给出了覆盖精确度和置信域的 Bartlett 修正, 把置信域的覆盖误差从 $O(n^{-1})$ 减少到 $O(n^{-2})$, 详细的讨论见 Cui 和 Chen (2003), 王启华等 (2010). Liu (2011) 研究了响应变量删失的线性 EV 模型的经验似然统计推断, 构造了模型中兴趣参数的经验似然置信域. 李高荣等 (2007, 2009) 分别研究了响应变量删失和缺失的线性 EV 模型, 借助核实数据, 发展了经验似然方法.

2.2.3 SIMEX 估计方法

Cook 和 Stefanski (1994) 针对参数测量误差模型, 假设测量误差的分布为正态分布时, 提出了 SIMEX (Simulation-Extrapolation) 方法对模型中参数估计的偏差进行修正. Carroll 等 (1996) 进一步研究了 SIMEX 估计的渐近性质. 自从 SIMEX 方法被提出以来, 已经成为研究测量误差模型的一个强有力的工具. 例如, Stefanski 和 Bay (1996), Carroll 等 (1999), Staudenmayer 和 Ruppert (2004), Liang 和 Ren (2005), Ma 和 Carroll (2006), Delaigle 和 Hall (2008), Liang 等 (2008), Apanasovich 和 Carroll (2009), Ma 和 Li (2010), Ma 和 Yin (2011), Sinha 和 Ma (2014).

假设 $\{(Y_i, W_i), i=1,\cdots,n\}$ 是来自于模型 (2.1.1) 和 (2.1.2) 的 i.i.d. 的随机样本, 并且假设 Σ_{uu} 已知. 当 Σ_{uu} 未知时, 可采用式 (2.2.3) 定义的重复测量的方法进行估计得到 $\hat{\Sigma}_{uu}$. 下面介绍 SIMEX 估计程序.

模拟步 (simulation step): 对于每一个 $i=1,\cdots,n$, 产生一系列变量

$$W_{ib}(\lambda) = W_i + (\lambda\Sigma_{uu})^{\frac{1}{2}} U_{ib}, \quad b=1,\cdots,B,$$

其中 $U_{ib} \sim N(0, I_p)$, I_p 是 $p\times p$ 的单位矩阵, B 是一个给定的任意大的正整数, $\lambda \in \Lambda = \{\lambda_1, \lambda_2, \cdots, \lambda_M\}$, 这里 Λ 表示 λ 取值的格子点, 一般 λ 的取值范围取为 [0,2].

估计步 (estimation step): 对所有的 $\lambda \in \Lambda = \{\lambda_1, \lambda_2, \cdots, \lambda_M\}$, 基于产生的新数据集 $\{(Y_i, W_{ib}(\lambda)), i=1,\cdots,n, b=1,\cdots,B\}$, 计算回归参数向量 β 的最小二乘估计如下

$$\hat{\beta}_b(\lambda) = \left(\frac{1}{n}\sum_{i=1}^n W_{ib}(\lambda)W_{ib}^{\mathrm{T}}(\lambda)\right)^{-1}\frac{1}{n}\sum_{i=1}^n W_{ib}(\lambda)Y_i, \quad b=1,\cdots,B.$$

进一步, 对所得估计 $\hat{\beta}_1(\lambda), \cdots, \hat{\beta}_B(\lambda)$ 进行平均, 即

$$\hat{\beta}(\lambda) = \frac{1}{B}\sum_{b=1}^B \hat{\beta}_b(\lambda).$$

外推步 (extrapolation step): 令 $\mathcal{G}(\lambda,\Gamma)$ 是一个外推函数, 并且令 $\hat{\beta}_j(\lambda)$ 是 $\hat{\beta}(\lambda)$ 的第 j 个分量, $j=1,\cdots,p$. 基于外推函数 $\mathcal{G}(\lambda,\Gamma)$, 对 $\{\hat{\beta}(\lambda),\lambda\in\Lambda\}$ 在 $\{\lambda\in\Lambda\}$ 上关于 Γ 作拟合回归模型如下

$$\sum_{\lambda\in\Lambda}[\hat{\beta}_j(\lambda)-\mathcal{G}(\lambda,\Gamma)]^2,\quad j=1,\cdots,p.$$

则 β_j 的 SIMEX 估计定义为

$$\hat{\beta}_{\mathrm{SIMEX},j}=\mathcal{G}(-1,\hat{\Gamma}),\quad j=1,\cdots,p.$$

如果 λ 退化到 0, $\hat{\beta}_j=\mathcal{G}(0,\hat{\Gamma})$ 就表示忽略测量误差后得到的估计, 即直接用 W 代替 X 得到的最小二乘估计.

注 2.2.1 在外推步中, 需要选取外推函数 $\mathcal{G}(\lambda,\Gamma)$, 通常取二次函数 $\mathcal{G}_Q(\lambda,\Gamma)=\gamma_1+\gamma_2\lambda+\gamma_3\lambda^2$, 或非线性函数 $\mathcal{G}_R(\lambda,\Gamma)=\gamma_1+\gamma_2/(\gamma_3+\lambda)$, 其中 $\Gamma=(\gamma_1,\gamma_2,\gamma_3)^{\mathrm{T}}$. 更多的讨论见 Cook 和 Stefanski (1994), Lin 和 Carroll (2000), Liang 和 Ren (2005), Liang 等 (2008).

2.3 响应变量缺失的线性测量误差模型

本节主要针对线性测量误差模型 (2.1.1) 和 (2.1.4), 响应变量随机缺失的情况下, 研究回归系数和响应均值的经验似然推断问题.

在实际中, 经常会遇到响应变量缺失的情况, 这时通常的统计推断方法不能直接应用. 通常处理缺失数据的方法是对每个缺失的响应变量值进行借补, 然后利用标准方法和借补后的完全数据进行研究. 对于线性模型, 在响应变量随机缺失和删失的情形, 利用 Owen (1991) 所提出的经验似然方法, Wang 和 Rao (2002b) 与 Qin 和 Jing (2001) 分别对这两种情况进行了研究. 而对协变量带有测量误差且响应变量缺失的情况, 本节借助于核实数据, 利用经验似然和借补方法研究了响应变量缺失而协变量带有误差的模型 (2.1.1), 构造了未知参数的两种经验对数似然比统计量, 即估计的经验对数似然比和调整的经验对数似然比. 假定随机缺失机制是 MAR(missing at random) 的, 即选择概率满足 $P(\delta=1|W,Y)=P(\delta=1|W)=:\pi(W)>0$. 此外, 基于借补方法也构造了响应变量均值的两种经验对数似然比统计量, 并证明了所提出的统计量渐近于 χ^2 分布.

2.3.1 回归系数的经验似然

假设主要数据 $\{(W_i,Y_i,\delta_i)_{i=1}^n\}$ 是 i.i.d. 的样本, 且与 i.i.d. 的核实样本 $\{(X_i,W_i)_{i=n+1}^{n+m}\}$ 相互独立, 其中当 Y_i 缺失时 $\delta_i=0$, 否则 $\delta_i=1$. 进一步假定 $E(\varepsilon|W)=0$.

令 $u(W)=E(X|W)$, 则基于完全观测数据, 模型 (2.1.1) 可改写为

$$\delta Y=\delta u^{\mathrm{T}}(W)\beta+\delta\eta, \tag{2.3.1}$$

其中 $\eta=\varepsilon+X^{\mathrm{T}}\beta-u^{\mathrm{T}}(W)\beta$. 现引入辅助随机变量

$$Z_i(\beta)=\delta_i u(W_i)(Y_i-u^{\mathrm{T}}(W_i)\beta). \tag{2.3.2}$$

由于 $u(W_i)$ 未知, 为了构造经验似然比, 首先需要给出 $u(W_i)$ 的估计. 利用核实数据, $u(W_i)$ 的估计定义为

$$\hat{u}_m(\tilde{x})=\left\{\sum_{j=n+1}^{n+m}X_jK\left(\frac{\tilde{x}-W_j}{h_m}\right)\right\}\bigg/\left\{\sum_{j=n+1}^{n+m}K\left(\frac{\tilde{x}-W_j}{h_m}\right)\right\}, \tag{2.3.3}$$

其中 $K(\cdot)$ 为核函数, h_m 是收敛于 0 的窗宽. 为了避免上述估计中由于分母的值太小而带来的技术困难, 需要将它进行改进. 为此, 记

$$\hat{f}_m(\tilde{x})=\frac{1}{mh_m^p}\sum_{j=n+1}^{n+m}K\left(\frac{\tilde{x}-W_j}{h_m}\right),$$

且 $\tilde{f}_m(\tilde{x})=\max\{\hat{f}_m(\tilde{x}),b_m\}$, 其中 b_m 是某个正的趋于 0 的常数列, 则将 $\hat{u}_m(\tilde{x})$ 重写为

$$\tilde{u}_m(\tilde{x})=\frac{\hat{u}_m(\tilde{x})\hat{f}_m(\tilde{x})}{\tilde{f}_m(\tilde{x})}. \tag{2.3.4}$$

用 $\tilde{u}_m(\tilde{x})$ 代替式 (2.3.2) 中的 $u(\tilde{x})$, 则记 $Z_{im}(\beta)=\delta_i\tilde{u}_m(W_i)(Y_i-\tilde{u}_m^{\mathrm{T}}(W_i)\beta)$. 假设 $p_1,\cdots,p_n$ 是一列非负数, 其和为 1, 则定义估计的经验对数似然比为

$$\hat{l}(\beta)=-2\max\left\{\sum_{i=1}^{n}\log(np_i)\;\middle|\;p_i\geqslant 0,\sum_{i=1}^{n}p_i=1,\sum_{i=1}^{n}p_iZ_{im}(\beta)=0\right\}. \tag{2.3.5}$$

由 Lagrange 乘子法, $\hat{l}(\beta)$ 可以表示为

$$\hat{l}(\beta)=2\sum_{i=1}^{n}\log\{1+\lambda^{\mathrm{T}}Z_{im}(\beta)\}, \tag{2.3.6}$$

其中 λ 满足

$$\frac{1}{n}\sum_{i=1}^{n}\frac{Z_{im}(\beta)}{1+\lambda^{\mathrm{T}}Z_{im}(\beta)}=0. \tag{2.3.7}$$

用 c 表示不依赖于 n 的正常数, 每次出现时可以取不同的值. 设 $\mathcal{D}^k$ 是 $\mathbb{R}^p$ 或其子区域上所有连续函数 φ 组成的类, 使得对 $0 < i_1 + \cdots + i_{p+1} \leqslant k, (k > p+1)$, 偏导数

$$\frac{\partial^{i_1}}{\partial x_1^{i_1}} \cdot \frac{\partial^{i_2}}{\partial x_2^{i_2}} \cdots \frac{\partial^{i_{p+1}}}{\partial x_{p+1}^{i_{p+1}}} \varphi(x_1, x_2, \cdots, x_{p+1})$$

一致有界. 对任何向量 a, 用 $\|a\|$ 表示 Euclidean 模. 为了得到 $\hat{l}(\beta)$ 的渐近分布, 需要下面的条件.

(C2.3.1) 假设 $E\|X\|^2 < \infty$ 和 $EY^2 < \infty$.

(C2.3.2) $\sup\limits_w E[\varepsilon^2|W=w] < \infty$.

(C2.3.3) $u(\cdot) \in \mathcal{D}^k,\ k > p+1$.

(C2.3.4) (i) W 的密度函数 $f_W(w) \in \mathcal{D}^k, k > p+1$.

(ii) 存在一个正的常数序列 b_m, 使得 $\limsup\limits_{m\to\infty} mP\{f_W(w) < b_m\} < \infty$.

(C2.3.5) $K(\cdot)$ 是非负有界的 k 阶核函数, 且具有有界支撑.

(C2.3.6) $mb_m^4 h_m^{2(p+1)} \to \infty,\quad mh_m^{2k} b_m^{-2} \to 0$.

(C2.3.7) $\Sigma = E[\delta u(W) u^{\mathrm{T}}(W)]$ 为正定阵.

(C2.3.8) $n/m \to r$, 其中 $r \geqslant 0$ 为常数.

上述条件是研究半参数模型中估计的渐近性质时常用的条件, 通常它们是可以满足的. 在此仅说明一下条件 (C2.3.6). 如果选取 $b_m = c_1 m^{\frac{p+1-k}{4(p+1+k)}} \log m$, 则取 $h_m = c_2 m^{-\frac{1}{p+1+k}}$, 其中 c_1 和 c_2 是正的常数.

定理 2.3.1 假设条件 (C2.3.1) ~ 条件 (C2.3.8) 成立. 如果 β 是参数真值, 当 $n \to \infty$ 时, 则有

$$\hat{l}(\beta) \xrightarrow{L} w_1 \chi_{1,1}^2 + \cdots + w_p \chi_{1,p}^2,$$

其中 "$\xrightarrow{L}$" 表示依分布收敛, $\chi_{1,i}^2 (1 \leqslant i \leqslant p)$ 是相互独立的 χ_1^2 变量, $w_i (1 \leqslant i \leqslant p)$ 是 $D(\beta) = \Lambda_0^{-1}(\beta)\Lambda(\beta)$ 的特征根, 并且

$$\begin{cases} \Lambda_0(\beta) = E[\delta u(W) u^{\mathrm{T}}(W) \{Y - u^{\mathrm{T}}(W)\beta\}^2], \\ \Lambda(\beta) = \Lambda_0(\beta) + rE(\delta u(W) u^{\mathrm{T}}(W) [\delta\{X - u(W)\}^{\mathrm{T}} \beta]^2). \end{cases}$$

为了应用定理 2.3.1 构造 β 的置信域或置信区间, 必须给出未知权 w_i 的相合估计. 由 plug-in 方法, $\Lambda_0(\beta)$ 和 $\Lambda(\beta)$ 可分别估计如下

$$\begin{cases} \hat{\Lambda}_0(\hat{\beta}) = \dfrac{1}{n} \displaystyle\sum_{i=1}^{n} [\delta_i \tilde{u}_m(W_i) \tilde{u}_m^{\mathrm{T}}(W_i) \{Y_i - \tilde{u}_m^{\mathrm{T}}(W_i)\hat{\beta}\}^2], \\ \hat{\Lambda}(\hat{\beta}) = \hat{\Lambda}_0(\hat{\beta}) + \dfrac{n}{m^2} \displaystyle\sum_{i=n+1}^{n+m} (\delta_i \tilde{u}_m(W_i) \tilde{u}_m^{\mathrm{T}}(W_i) [\delta_i \{X_i - \tilde{u}_m(W_i)\}^{\mathrm{T}} \hat{\beta}]^2). \end{cases}$$

此处 $\hat{\beta}$ 是基于式 (2.3.1) 定义的 β 的最小二乘估计, 即

$$\hat{\beta}=\left[\frac{1}{n}\sum_{i=1}^{n}\delta_i\tilde{u}_m(W_i)\tilde{u}_m^{\mathrm{T}}(W_i)\right]^{-1}\left[\frac{1}{n}\sum_{i=1}^{n}\delta_i\tilde{u}_m^{\mathrm{T}}(W_i)Y_i\right]. \tag{2.3.8}$$

由此可以推出 $\hat{D}(\hat{\beta})=\hat{\Lambda}_0^{-1}(\hat{\beta})\hat{\Lambda}(\hat{\beta})$ 的特征根 $\hat{w}_i$ 是 $w_i(i=1,\cdots,p)$ 的相合估计. 记 $\hat{S}=\hat{w}_1\chi_{1,1}^2+\cdots+\hat{w}_p\chi_{1,p}^2$. 用 $H(\cdot)$ 表示在给定 $\{(W_i,Y_i,\delta_i)_{i=1}^n\}$ 和 $\{(X_i,W_i)_{i=n+1}^{n+m}\}$ 下 $\hat{S}$ 的条件分布. 设 $\hat{c}_\alpha$ 是 $H(\cdot)$ 的 $1-\alpha$ 分位点, 则 β 的置信度为 $1-\alpha$ 的置信域为

$$\hat{I}_\alpha(\tilde{\beta})=\{\tilde{\beta}\in\mathbb{R}^p \mid \hat{l}(\tilde{\beta})\leqslant\hat{c}_\alpha\}. \tag{2.3.9}$$

实际上, 要得到分布 $H(\cdot)$ 是很方便的. 它可以通过重复产生 χ_1^2 样本 $\chi_{1,1}^2,\cdots,\chi_{1,p}^2$, 然后使用 Monte Carlo 模拟而得到.

注 2.3.1 当 $\pi(W)\equiv 1$, 即 Y 完全观测时, 定理 2.3.1 就是 Wang 和 Rao (2002a) 的结果. 此外, 如果协变量也不存在测量误差时, $D(\beta)$ 为单位阵, 则 $w_i=1(i=1,\cdots,p)$, 这时定理 2.3.1 便为 Owen (1991) 的结果, 即经验对数似然比函数渐近于自由度为 p 的 χ^2 分布.

如前所述, 在应用定理 2.3.1 构造参数 β 的置信域, 需要估计权 w_i, 这就会降低置信域的精度, 因此需要对经验似然比进行调整. 按照 Rao 和 Scott (1981) 的结果, $\rho(\beta)\sum_{i=1}^{p}w_i\chi_{1,i}^2$ 的分布可用自由度为 p 的 χ^2 分布逼近, 这里 $\rho(\beta)=p/\mathrm{tr}\{D(\beta)\}$. 由定理 2.3.1 以及 $\hat{\Lambda}(\hat{\beta})$ 和 $\hat{\Lambda}_0(\hat{\beta})$ 的相合性, 可以推出 $\hat{\rho}(\hat{\beta})\hat{l}(\beta)$ 渐近于 χ^2 分布, 此处 $\hat{\rho}(\hat{\beta})=p/\mathrm{tr}\{\hat{D}(\hat{\beta})\}$. 为了提高逼近精度, 在 $\hat{\rho}(\hat{\beta})$ 中用 β 代替 $\hat{\beta}$. 因此, 改良 Rao 和 Scott 调整的经验对数似然可定义为 $\tilde{l}(\beta)=\hat{\rho}(\beta)\hat{l}(\beta)$. 然而, 这种逼近的精度依赖于 w_i 的值. 下面给出一个调整的经验对数似然比. 注意到

$$\hat{\rho}(\beta)=\frac{\mathrm{tr}\{\hat{\Lambda}^{-1}(\beta)\hat{\Lambda}(\beta)\}}{\mathrm{tr}\{\hat{\Lambda}_0^{-1}(\beta)\hat{\Lambda}(\beta)\}}.$$

检查 $\hat{l}(\beta)$ 渐近表达式, 在 $\hat{\rho}(\beta)$ 中用 $\hat{B}(\beta)=\left\{\sum_{i=1}^{n}Z_{im}(\beta)\right\}\left\{\sum_{i=1}^{n}Z_{im}(\beta)\right\}^{\mathrm{T}}$ 代替$\hat{\Lambda}(\beta)$, 就得到一个不同的调整因子 $\hat{A}(\beta)=\dfrac{\mathrm{tr}\{\hat{\Lambda}^{-1}(\beta)\hat{B}(\beta)\}}{\mathrm{tr}\{\hat{\Lambda}_0^{-1}(\beta)\hat{B}(\beta)\}}$. 因此定义调整的经验对数似然比为

$$\hat{l}_{\mathrm{ad}}(\beta)=\hat{A}(\beta)\hat{l}(\beta).$$

定理 2.3.2 假设条件 (C2.3.1) ~条件 (C2.3.8) 成立. 如果 β 是真实参数, 当 $n\to\infty$ 时, 则有

$$\hat{l}_{\mathrm{ad}}(\beta)\xrightarrow{L}\chi_p^2.$$

基于定理 2.3.2, 可以使用 $\hat{l}_{\mathrm{ad}}(\beta)$ 构造 β 的置信域

$$\hat{I}_{\mathrm{ad},\alpha}(\tilde{\beta}) = \{\tilde{\beta} \in \mathbb{R}^p \mid \hat{l}_{\mathrm{ad}}(\tilde{\beta}) \leqslant c_\alpha\}, \tag{2.3.10}$$

这里 $P(\chi_p^2 \leqslant c_\alpha) = 1 - \alpha$. 并且有 $P(\beta \in \hat{I}_{\mathrm{ad},\alpha}(\tilde{\beta})) = 1 - \alpha + o(1)$.

为了能够将经验似然方法与最小二乘方法进行比较, 在此给出 $\hat{\beta}$ 的渐近正态性, 其中 $\hat{\beta}$ 是式 (2.3.8) 所定义的最小二乘估计.

定理 2.3.3 假设条件 (C2.3.1) ~ 条件 (C2.3.8) 成立, 当 $n \to \infty$ 时, 则有

$$\sqrt{n}(\hat{\beta} - \beta) \xrightarrow{L} N(0, \Sigma^{-1}\Lambda(\beta)\Sigma^{-1}),$$

其中 Σ 在条件 (C2.3.7) 中定义, 而 $\Lambda(\beta)$ 在定理 2.3.1 中定义. 记

$$\hat{\Sigma} = \frac{1}{n}\sum_{i=1}^{n} \delta_i \tilde{u}_m(W_i)\tilde{u}_m^{\mathrm{T}}(W_i),$$

类似于 Wang (1999) 中引理 A.2 可证得 $\hat{\Sigma} \xrightarrow{P} \Sigma$.

2.3.2 响应均值的经验似然

本节考虑响应均值 (记为 θ) 的统计推断问题, 利用经验似然方法构造了 θ 的一个调整的经验对数似然比函数, 并证明了该函数渐近收敛到一个 χ_1^2 分布. 由 $u(\tilde{x})$ 和 β 分别在式 (2.3.4) 和式 (2.3.8) 所定义的估计 $\tilde{u}_m(W_i)$ 和 $\hat{\beta}$ 可得 Y_i 的一个回归借补, 令

$$\widetilde{Y}_{in} = \delta_i Y_i + (1 - \delta_i)\tilde{u}_m^{\mathrm{T}}(W_i)\hat{\beta}, \quad i = 1, \cdots, n. \tag{2.3.11}$$

则当 $\theta = \theta_0$ 为真参数时, 有 $E\widetilde{Y}_{in} = EY_i = \theta_0$. 因此, $\widetilde{Y}_{in}$ 可以看成是 θ 的估计, 记为 $\hat{\theta}$. 假设 $q_1, \cdots, q_n$ 是一列非负数, 其和为 1, 则定义估计的经验对数似然比为

$$\hat{R}_n(\theta_0) = -2\max\left\{\sum_{i=1}^{n}\log(nq_i) \,\middle|\, q_i \geqslant 0, \sum_{i=1}^{n} q_i = 1, \sum_{i=1}^{n} q_i\widetilde{Y}_{in} = \theta_0\right\}. \tag{2.3.12}$$

由 Lagrange 乘子法, $\hat{R}_n(\theta_0)$ 可以表示为

$$\hat{R}_n(\theta_0) = 2\sum_{i=1}^{n}\log\{1 + \gamma(\widetilde{Y}_{in} - \theta_0)\}, \tag{2.3.13}$$

其中 γ 满足

$$\frac{1}{n}\sum_{i=1}^{n}\frac{\widetilde{Y}_{in} - \theta_0}{1 + \gamma(\widetilde{Y}_{in} - \theta_0)} = 0.$$

与 Owen (1991) 的经验对数似然比函数相比, 主要区别是 $\widetilde{Y}_{in}$ 不是 i.i.d.. 因此, $\hat{R}_n(\theta_0)$ 的渐近分布不是一个 χ_1^2 分布. 为了得到本节的主要结果, 给出下面两个关键的引理. 引理 2.3.1 和引理 2.3.2 的证明细节见李高荣等 (2009).

引理 2.3.1 假设条件 (C2.3.1) ~ 条件 (C2.3.8) 成立, 如果 θ_0 是 θ 的真值, 当 $n \to \infty$ 时, 则有

$$\frac{1}{\sqrt{n}}\sum_{i=1}^{n}(\widetilde{Y}_{in}-\theta_0) \xrightarrow{L} N(0,\Xi(\theta_0)),$$

其中

$$\Xi(\theta_0)=\Sigma_1+\beta^{\mathrm{T}}\Sigma_3\beta-2\Sigma_4^{\mathrm{T}}\beta\theta_0+\theta_0^2+\Sigma_2^{\mathrm{T}}(\Sigma^{-1}\Lambda(\beta)\Sigma^{-1})\Sigma_2+\Sigma_2^{\mathrm{T}}(\Sigma^{-1}\Lambda(\beta)\Sigma^{-1})\Sigma_5,$$

且

$$\Sigma_1=E[\delta(Y-u^{\mathrm{T}}(W)\beta)^2],\quad \Sigma_2=E[(1-\delta)u(W)],\quad \Sigma_3=E[u(W)u^{\mathrm{T}}(W)],$$
$$\Sigma_4=E[u(W)],\quad \Sigma_5=E[\delta u(W)],$$

Σ 在条件 (C2.3.7) 中定义, $\Lambda(\beta)$ 在定理 2.3.1 中定义.

引理 2.3.2 假设条件 (C2.3.1) ~ 条件 (C2.3.8) 成立, 当 $n \to \infty$ 时, 则有

$$\tilde{\Xi}(\theta)=\frac{1}{n}\sum_{i=1}^{n}(\widetilde{Y}_{in}-\theta_0)^2=\Sigma_1-2\Sigma_4^{\mathrm{T}}\beta\theta_0+\beta^{\mathrm{T}}\Sigma_3\beta+\theta_0^2+o_P(1).$$

从引理 2.3.1 和引理 2.3.2 可以推出, $\frac{1}{n}\sum_{i=1}^{n}(\widetilde{Y}_{in}-\theta_0)^2$ 依概率不能收敛到

$$\frac{1}{\sqrt{n}}\sum_{i=1}^{n}(\widetilde{Y}_{in}-\theta_0)$$

的渐近方差.

下面获得调整的经验对数似然比 $\hat{R}_{\mathrm{ad}}(\theta)$. 令

$$\tilde{\Xi}_n(\theta)=\frac{1}{n}\sum_{i=1}^{n}(\widetilde{Y}_{in}-\theta)^2,$$

和在引理 2.3.1 中定义的方差 $\Xi(\theta)$ 的估计 $\hat{\Xi}_n(\theta)$ 为

$$\begin{aligned}\hat{\Xi}_n(\theta)=&\Sigma_{1n}+\hat{\beta}^{\mathrm{T}}\Sigma_{3n}\hat{\beta}-2\Sigma_{4n}^{\mathrm{T}}\hat{\beta}\theta+\theta^2+\Sigma_{2n}^{\mathrm{T}}\left(\hat{\Sigma}^{-1}\hat{\Lambda}(\hat{\beta})\hat{\Sigma}^{-1}\right)\Sigma_{2n}\\&+\Sigma_{2n}^{\mathrm{T}}\left(\hat{\Sigma}^{-1}\hat{\Lambda}(\hat{\beta})\hat{\Sigma}^{-1}\right)\Sigma_{5n},\end{aligned}$$

其中 $\hat{\Sigma}$ 在定理 2.3.3 中定义, $\hat{\Lambda}(\hat{\beta})$ 在定理 2.3.1 中定义, 且

$$\Sigma_{1n}=\frac{1}{n}\sum_{i=1}^{n}\delta_i(Y_i-\tilde{u}_m^{\mathrm{T}}(W_i)\hat{\beta})^2,\quad \Sigma_{2n}=\frac{1}{n}\sum_{i=1}^{n}(1-\delta_i)\tilde{u}_m(W_i),$$
$$\Sigma_{3n}=\frac{1}{n}\sum_{i=1}^{n}\tilde{u}_m(W_i)\tilde{u}_m^{\mathrm{T}}(W_i),\quad \Sigma_{4n}=\frac{1}{n}\sum_{i=1}^{n}\tilde{u}_m(W_i),\quad \Sigma_{5n}=\frac{1}{n}\sum_{i=1}^{n}\delta_i\tilde{u}_m(W_i).$$

令 $r_n(\theta)=\dfrac{\tilde{\Xi}_n(\theta)}{\hat{\Xi}_n(\theta)}$, 则定义了下面的调整的经验对数似然比函数为

$$\hat{R}_{\mathrm{ad}}(\theta)=r_n(\theta)\hat{R}_n(\theta). \tag{2.3.14}$$

定理 2.3.4 假设条件 (C2.3.1) ~条件 (C2.3.8) 成立. 如果 θ_0 是 θ 的真值, 当 $n\to\infty$ 时, 则有

$$\hat{R}_{\mathrm{ad}}(\theta_0)\xrightarrow{L}\chi_1^2.$$

基于定理 2.3.4, 可以使用 $\hat{R}_{\mathrm{ad}}(\theta)$ 构造 θ 的置信域

$$\hat{I}_{\mathrm{ad},\alpha}=\{\tilde{\theta}\in\mathbb{R}\mid\hat{R}_{\mathrm{ad}}(\tilde{\theta})\leqslant c_\alpha\}, \tag{2.3.15}$$

这里 $P(\chi_1^2\leqslant c_\alpha)=1-\alpha$, 并且有 $P(\theta\in\hat{I}_{\mathrm{ad},\alpha})=1-\alpha+o(1)$.

注 2.3.2 当协变量不带有测量误差时, 即 X 能准确测量时, 定理 2.3.4 就是 Wang 和 Rao (2002b) 中定理 1 的结果. 明显 $r_n(\theta)\xrightarrow{P}r(\theta)$, 其中 $r(\theta)=\dfrac{\tilde{\Xi}(\theta)}{\Xi(\theta)}$, $\Xi(\theta)$ 和 $\tilde{\Xi}(\theta)$ 分别在引理 2.3.1 和引理 2.3.2 中定义. 此外, 当 $\pi(W)\equiv 1$, 即 Y 完全观测时, 有 $r(\theta)=1$, 则所定义的调整的经验对数似然函数就成为标准的经验对数似然函数.

为了能够将经验似然方法与基于渐近正态结果进行比较, 下面给出 $\hat{\theta}$ 的渐近正态性.

定理 2.3.5 假设条件 (C2.3.1) ~ 条件 (C2.3.8) 成立, 当 $n\to\infty$ 时, 则有

$$\sqrt{n}(\hat{\theta}-\theta)\xrightarrow{L}N(0,\Xi(\theta)),$$

其中 $\Xi(\theta)$ 在引理 2.3.1 中定义, 其估计为 $\hat{\Xi}_n(\theta)$.

2.3.3 模拟研究

本小节通过模拟研究比较三种方法得到的回归系数 β 的置信区间的覆盖概率大小及其区间平均长度大小. 它们分别是: 估计的经验似然 (EEL)、调整的经验似然 (AEL) 和基于最小二乘 (LS) 估计所得的渐近正态分布. 此外, 还用两种方法对响应变量均值 θ 的置信区间的覆盖概率大小及其区间平均长度进行了比较. 这两种方法是: 调整的经验似然 (AEL) 和基于定理 2.3.5 所得的渐近正态分布 (NA). 这里考虑选择概率分别取

$$\pi_1=P(\delta=1|W)=0.6,\quad \pi_2=P(\delta=1|W)=0.8.$$

核实样本和主要样本大小分别取 $(m,n)=(10,30),(30,90),(60,180),(10,50),(30,150)$ 和 $(60,300)$.

考虑模型: $Y = X\beta+\varepsilon$, $W = 1.25X^2+U$, 其中测量误差 U 取 $[-1,1]$ 上的均匀分布. 这里 $\beta = 2$, $X \sim N(2,1)$, $\varepsilon \sim N(0,1)$. 取核函数 $K(u) = \frac{1}{2}(3-u^2)\frac{1}{\sqrt{2\pi}}\mathrm{e}^{-\frac{u^2}{2}}$, 其中 $|u| < \sqrt{3}$. 窗宽和截断常数的选取是: 首先取定一个尽可能小的 b_m 作为初始值, 然后利用交错核实方法选一个最优窗宽 $h_{\rm opt}$, 最后确定一个 $b_{\rm opt}$, 使得 $\tilde{u}_m(\tilde{x})$ 的均方误差达到最小. 最优窗宽的速度是 $m^{-\frac{1}{p+1+2k}}$, 而条件中只能取到窗宽的速度是 $m^{-\frac{1}{p+1+k}}$. 最后取定的窗宽是 $h_m = h_{\rm opt}m^{-\frac{k}{(p+1+k)(p+1+2k)}}$, 满足条件要求. 置信度 $1-\alpha = 0.95$, 做 1000 次重复运算. 模拟结果分别在表 2.3.1 和表 2.3.2 中给出.

表 2.3.1　β 的置信度为 95%的置信区间的覆盖概率和平均区间长度

$\pi(x)$	$r=\frac{n}{m}$	(m,n)	覆盖概率			平均区间长度		
			EEL	AEL	LS	EEL	AEL	LS
		(10,30)	0.8380	0.8460	0.9720	0.9368	0.9053	1.5420
π_1	3	(30,90)	0.8850	0.9260	0.9560	0.5517	0.5135	1.0440
		(60,180)	0.9240	0.9320	0.9750	0.4726	0.4575	0.8861
		(10,30)	0.8670	0.9090	0.9690	0.9179	0.8944	1.4670
π_2	3	(30,90)	0.9183	0.9330	0.9670	0.5291	0.4963	0.8674
		(60,180)	0.9370	0.9410	0.9450	0.4482	0.4239	0.7802
		(10,50)	0.8810	0.9040	0.9580	0.9225	0.8981	1.3630
π_1	5	(30,150)	0.9110	0.9290	0.9660	0.4976	0.4823	0.8936
		(60,300)	0.9360	0.9410	0.9550	0.4114	0.3918	0.6629
		(10,50)	0.8920	0.9190	0.9640	0.8880	0.8735	1.3070
π_2	5	(30,150)	0.9230	0.9380	0.9290	0.4815	0.4653	0.6209
		(60,300)	0.9390	0.9430	0.9430	0.4028	0.3832	0.5130

从表 2.3.1 可以看出, 调整的经验似然 (AEL) 明显比其他两种方法好, 它的置信区间的平均区间长度比 EEL 和 LS 都小, 而覆盖概率随着样本容量的增加接近于置信度 95%. 虽然随着核实样本和主要数据样本大小 m,n 的增大, 这三种方法所得的覆盖概率都接近 95%. 但是 LS 的覆盖概率最大, 并且超过了置信度 95%, 同时它有较长的平均区间长度, 并注意到 LS 的平均区间长度几乎是 EEL 和 AEL 的 1.5 倍左右, 这是因为最小二乘方法需要估计渐近方差所致. EEL 和 AEL 的比较可以看出 AEL 模拟结果明显要好. 另外当选择概率大时, 即缺失概率小时, 这三种方法所得的置信区间的精确度要比选择概率小时要好.

从表 2.3.2 对响应变量均值 θ 的模拟结果可以看出, AEL 明显比 NA 方法好. 因为 AEL 比 NA 的覆盖概率大, 而平均区间长度短. 随着核实样本和主要数据样本大小 m,n 的增大, AEL 和 NA 的覆盖概率都接近 95%. 而且当选择概率大时, AEL 和 NA 的置信区间的精确度要比选择概率小时要好.

表 2.3.2 θ 的置信度为 95%的置信区间的覆盖概率和平均区间长度

$\pi(x)$	$r=\frac{n}{m}$	(m,n)	覆盖概率		平均区间长度	
			AEL	NA	AEL	NA
π_1	3	(10,30)	0.9120	0.9030	1.1085	1.4172
		(30,90)	0.9230	0.9110	0.7656	1.1458
		(60,180)	0.9400	0.9370	0.7139	1.0265
π_2	3	(10,30)	0.9140	0.9090	0.9349	1.2530
		(30,90)	0.9270	0.9150	0.7387	1.0211
		(60,180)	0.9460	0.9420	0.6993	0.8944
π_1	5	(10,50)	0.9180	0.9130	0.9296	1.2515
		(30,150)	0.9260	0.9210	0.7255	1.0665
		(60,300)	0.9450	0.9420	0.6804	0.8231
π_2	5	(10,50)	0.9220	0.9190	0.9020	1.1743
		(30,150)	0.9340	0.9320	0.6862	0.9497
		(60,300)	0.9480	0.9450	0.6721	0.7957

综合表 2.3.1 和表 2.3.2 得出如下结论: 调整的经验似然 (AEL) 是一个优良的方法, 它得到的回归系数 β 和响应变量均值 θ 的置信域具有较高的精确度和置信度. 此外, 也对线性测量误差的情况进行了模拟计算, 其模拟效果更好一些.

2.4 广义线性测量误差模型的纠偏经验似然方法

本节的主要目的是针对逆高斯回归、泊松 (Poisson) 回归和 Logistic 回归等一类广义线性测量误差模型 (2.1.5) 和 (2.1.6), 构造模型中 β 的置信域. 为了消除测量误差对参数估计所引起的影响, 基于纠偏的得分函数的矩可识别性, 发展了纠偏的经验似然方法研究一类广义线性测量误差模型的统计推断问题. 在一定的正则条件下, 证明了回归参数的纠偏经验对数似然比统计量渐近收敛到标准 χ^2 分布, 同时证明了回归参数极大经验似然估计的渐近正态性.

为了构造 β 的置信域, 最直接的方法是使用 β 估计量的渐近正态分布. 使用渐近正态分布构造 β 置信域时, 首先需要使用 plug-in 方法估计渐近方差矩阵, 而渐近方差矩阵的 plug-in 估计的复杂性和不稳定性会引起 β 置信域构造较大的误差. β 置信域构造的另一个方法是 bootstrap 方法, 主要优点是置信域不依赖于 β 估计量的渐近分布, 而难点是 bootstrap 方法的渐近性质很难研究, 特别对于测量误差情形. 考虑到这些情形, 此处推荐使用经验似然方法来构造 β 的置信域. 经验似然方法被 Owen (1988) 应用于线性回归模型的研究, Kolaczyk (1994) 进一步推广到广义线性模型的研究. Cui 和 Chen (2003) 与 Cui 和 Kong (2006) 分别把经验似然方法应用到线性测量误差模型和部分线性测量误差模型的研究. Stute 等 (2007) 研究了具有核实数据的非线性测量误差模型的经验似然推断问题. 更多相关的工作见

Wang 和 Rao (2002a), Xue (2009a, 2009b), Xue 和 Zhu (2007), Zhu 和 Xue (2006), Li 等 (2012a, 2010b) 等.

纠偏的经验似然方法有许多重要的性质: ① 为了消除测量误差对参数估计的影响, 基于 Nakamura (1990) 与 Huang 和 Wang (2001) 的矩可识别性的想法, 发展了纠偏的得分函数, 进而证明纠偏的经验对数似然比统计量渐近收敛到标准 χ^2 分布; ② 避免了估计渐近正态分布中的渐近方差, 这样有效地提高了经验似然置信域的精度和可靠度; ③ 经验似然置信域的形状和方向完全由数据来确定.

2.4.1 纠偏经验似然方法和主要结果

1. 纠偏得分函数

令 $\{(X_1, Y_1), \cdots, (X_n, Y_n)\}$ 是来自模型 (2.1.5) 的一组随机样本. 如果 X_i 可观测, β 的无偏得分函数为

$$\eta_i(\beta; X_i, Y_i) = \frac{\partial \mu_i}{\partial \beta} V_i^{-1}(\mu, \phi)(Y_i - \mu_i). \tag{2.4.1}$$

当 X_i 不能准确测量, 而观测到的是具有测量误差的变量 W_i 时, 直接用观测变量 W_i 代替变量 X_i, 则导致 $\eta_i(\beta; W_i, Y_i)$ 是一个有偏的得分函数, 即 $E\left[\eta_i(\beta; W_i, Y_i)\right] \neq 0$ 成立. 由 Nakamura (1990) 的想法, 构造 β 的无偏得分函数 $\eta_i^*(\Sigma_{uu}, \beta; W_i, Y_i)$, 使得

$$E\left[\eta_i^*(\Sigma_{uu}, \beta; W_i, Y_i)\right] = 0.$$

为了找到无偏得分函数, 由误差模型 (2.1.6) 和矩可识别性条件, 有

$$E(W|X) = X, \quad E(WW^{\mathrm{T}}|X) = XX^{\mathrm{T}} + \Sigma_{uu}, \tag{2.4.2}$$

$$E\left[W\exp(W^{\mathrm{T}}\beta)|X\right] = (X + \Sigma_{uu}\beta)\exp\left(X^{\mathrm{T}}\beta + \beta^{\mathrm{T}}\Sigma_{uu}\beta/2\right), \tag{2.4.3}$$

$$E\left[W\exp(-W^{\mathrm{T}}\beta)|X\right] = (X - \Sigma_{uu}\beta)\exp\left[-X^{\mathrm{T}}\beta + \beta^{\mathrm{T}}\Sigma_{uu}\beta/2\right], \tag{2.4.4}$$

$$E\left[W\exp(-2W^{\mathrm{T}}\beta)|X\right] = (X - 2\Sigma_{uu}\beta)\exp\left[-2X^{\mathrm{T}}\beta + 2\beta^{\mathrm{T}}\Sigma_{uu}\beta\right]. \tag{2.4.5}$$

下面就实际中经常使用的几种广义线性测量误差模型, 构造相应的纠偏得分函数.

(1) **伽马测量误差回归模型**　令 Y 来自于伽马分布, 密度函数为

$$f(y) = \frac{1}{\Gamma(\phi)\theta^{\phi}} y^{\phi-1}\exp(-y/\theta),$$

其中 ϕ 是已知的, θ 是典范参数, 且 $\Gamma(\cdot)$ 是伽马函数. 给定 X 条件下, Y 的均值和方差分别为 $\mu=\phi\theta$ 和 $\mathrm{Var}(Y|X)=\mu^2/\phi$. 考虑如下对数线性测量误差模型:

$$\begin{cases}\log(\mu_i)=X_i^{\mathrm{T}}\beta,\\ W_i=X_i+U_i.\end{cases}$$

由式 (2.4.1), 得分函数定义为

$$\eta_i(\beta;X_i,Y_i)=\phi X_i[Y_i\exp(-X_i^{\mathrm{T}}\beta)-1].$$

由矩可识别性条件 (2.4.2) 和 (2.4.4), 则有下面的纠偏的得分函数:

$$\eta_i^*(\Sigma_{uu},\beta;W_i,Y_i)=\phi\,(W_i+\Sigma_{uu}\beta)\exp(-W_i^{\mathrm{T}}\beta-\beta^{\mathrm{T}}\Sigma_{uu}\beta/2)Y_i-\phi W_i.$$

(2) **逆高斯或 Wald 测量误差回归模型** 令 Y 来自具有均值 μ 和方差 $\mathrm{Var}(Y|X)=\phi\mu^3$ 的逆高斯分布. 考虑对数线性测量误差模型:

$$\begin{cases}\log(\mu_i)=X_i^{\mathrm{T}}\beta,\\ W_i=X_i+U_i.\end{cases}$$

由矩可识别性条件 (2.4.4) 和 (2.4.5), 纠偏的得分函数为

$$\begin{aligned}\eta_i^*(\Sigma_{uu},\beta;W_i,Y_i)=&\,\phi^{-1}(W_i+2\Sigma_{uu}\beta)\exp\left(-2W_i^{\mathrm{T}}\beta-2\beta^{\mathrm{T}}\Sigma_{uu}\beta\right)Y_i\\&-\phi^{-1}(W_i+\Sigma_{uu}\beta)\exp\left(-W_i^{\mathrm{T}}\beta-\beta^{\mathrm{T}}\Sigma_{uu}\beta/2\right).\end{aligned}$$

(3) **泊松测量误差回归模型** 令 Y 来自均值为 μ 的泊松分布, 则 $\mathrm{Var}(Y|X)=\mu$. 考虑对数线性测量误差模型:

$$\begin{cases}\log(\mu_i)=X_i^{\mathrm{T}}\beta,\\ W_i=X_i+U_i.\end{cases}$$

容易推得纠偏的得分函数为

$$\eta_i^*(\Sigma_{uu},\beta;W_i,Y_i)=W_iY_i-(W_i-\Sigma_{uu}\beta)\exp\left(W_i^{\mathrm{T}}\beta-\beta^{\mathrm{T}}\Sigma_{uu}\beta/2\right).$$

(4) **二元 Logistic 测量误差回归模型** 考虑如下 Logistic 测量误差回归模型:

$$\begin{cases}P(Y=1|X)=\dfrac{1}{1+\exp(-X^{\mathrm{T}}\beta)},\\ W=X+U,\end{cases}$$

其中均值 $\mu=[1+\exp(-X^{\mathrm{T}}\beta)]^{-1}$, 方差 $\mathrm{Var}(Y|X)=\mu(1-\mu)$. 得分函数为

$$\eta_i(\beta;X_i,Y_i)=X_i\{Y_i-[1+\exp(-X_i^{\mathrm{T}}\beta)]^{-1}\}.$$

由 Nakamura (1990) 的想法, $X_i[1+\exp(-X_i^{\mathrm{T}}\beta)]^{-1}$ 不能被纠偏, 因此该模型不存在纠偏的得分函数. 下面由 Huang 和 Wang (2001) 的想法, 并对得分函数 $\eta_i(\beta;X_i,Y_i)$ 引入适当的权使得 $\eta_{i,\omega}(\beta;W_i,Y_i)$ 能够纠偏. 定义

$$\eta_{i,\omega}(\beta;X_i,Y_i)=\omega(\beta;X_i,Y_i)\eta_i(\beta;X_i,Y_i),$$

其中 $\omega(\beta;X_i,Y_i)=1+\exp(X_i^{\mathrm{T}}\beta)$. 如果存在一个函数 $\eta_i^*(\Sigma_{uu},\beta;W_i,Y_i)$ 使得

$$E\left[\eta_i^*(\Sigma_{uu},\beta;W_i,Y_i)\,|X_i,Y_i\right]=\eta_{i,\omega}(\beta;X_i,Y_i),$$

则 $\eta_i^*(\Sigma_{uu},\beta;W_i,Y_i)$ 是一个无偏得分函数. 由矩可识别性条件 (2.4.2) 和 (2.4.3), 纠偏的得分函数定义为

$$\eta_i^*(\Sigma_{uu},\beta;W_i,Y_i)=W_iY_i+(W_i+\Sigma_{uu}\beta)\exp\left(-W_i^{\mathrm{T}}\beta-\beta^{\mathrm{T}}\Sigma_{uu}\beta/2\right)Y_i-W_i.$$

2. 纠偏的经验似然方法和渐近性质

注意到$\left\{\eta_i^*(\Sigma_{uu},\beta;W_i,Y_i);1\leqslant i\leqslant n\right\}$ 是相互独立的, 且有 $E\{\eta_i^*(\Sigma_{uu},\beta;W_i,Y_i)\}=0$. 回归参数 β 的纠偏的经验对数似然比函数定义为

$$l(\Sigma_{uu},\beta)=-2\max\left\{\sum_{i=1}^n\log(np_i)\;\middle|\;p_i\geqslant 0,\sum_{i=1}^n p_i=1,\sum_{i=1}^n p_i\eta_i^*(\Sigma_{uu},\beta;W_i,Y_i)=0\right\},$$

其中 $p_i,i=1,\cdots,n$ 是非负实数. 由 Lagrange 乘子方法, 能够推得

$$l(\Sigma_{uu},\beta)=2\sum_{i=1}^n\log\left[1+\lambda^{\mathrm{T}}\eta_i^*(\Sigma_{uu},\beta;W_i,Y_i)\right],\tag{2.4.6}$$

其中 λ 是 $p\times 1$ 维的 Lagrange 乘子向量, 并是下面方程的解

$$\sum_{i=1}^n\frac{\eta_i^*(\Sigma_{uu},\beta;W_i,Y_i)}{1+\lambda^{\mathrm{T}}\eta_i^*(\Sigma_{uu},\beta;W_i,Y_i)}=0.\tag{2.4.7}$$

令$\Omega=E\left\{\eta^*(\Sigma_{uu},\beta;W,Y)\,\eta^{*\mathrm{T}}(\Sigma_{uu},\beta;W,Y)\right\}$. 假设参数空间 Θ 是紧的. 纠偏的经验对数似然比函数 $l(\Sigma_{uu},\beta)$ 的渐近分布的结果见定理 2.4.1.

定理 2.4.1　假设矩阵 Ω 是非奇异的, $E(XX^{\mathrm{T}})<\infty$ 和 $E(UU^{\mathrm{T}})<\infty$. 如果 β 是真实的回归参数, 且当 $n\to\infty$ 时, 则有

$$l(\Sigma_{uu},\beta)\xrightarrow{L}\chi_p^2,$$

其中 χ_p^2 表示自由度为 p 的 χ^2 分布.

对于 $0 < \alpha < 1$, 令 $\chi_p^2(\alpha)$ 表示 χ_p^2 分布的 α 分位点. 由定理 2.4.1, 回归参数 β 的 $1-\alpha$ 置信域定义为

$$I_\alpha(\breve{\beta}) = \left\{\breve{\beta} \in \mathbb{R}^p \;\middle|\; l(\breve{\beta}) \leqslant \chi_p^2(1-\alpha)\right\}.$$

注 2.4.1 因为直接用观测变量 W_i 代替变量 X_i 后得到的得分函数 $\eta_i(\beta; W_i, Y_i)$ 是有偏的得分函数, 因此容易证明忽略测量误差的自然的经验对数似然比 (naive empirical log-likelihood ratio, NELR) 统计量不能渐近收敛到一个标准 χ^2 分布.

极大化 $-l(\beta)$, 得到 β 的极大经验似然估计, 记为 $\hat{\beta}$. 当约束的个数等于参数的个数时, 可以证明最优的 $p_i = 1/n$ (Owen, 1988; Qin and Lawless, 1994), 则 β 的极大经验似然估计等价于下面纠偏估计方程的解

$$\sum_{i=1}^n \eta_i^* \left(\Sigma_{uu}, \beta; W_i, Y_i\right) = 0.$$

下面定理证明了 β 极大经验似然估计的渐近正态性.

定理 2.4.2 在定理 2.4.1 的条件下, 如果 Γ 是一个正定矩阵, 当 $n \to \infty$ 时, 则有

$$\sqrt{n}(\hat{\beta} - \beta) \xrightarrow{L} N\left(0, \Gamma^{-1}\Omega(\Gamma^{-1})^{\mathrm{T}}\right),$$

其中 $\Gamma = E\left\{\partial\eta^*\left(\Sigma_{uu}, \beta; W, Y\right)/\partial\beta\right\}$.

为了使用定理 2.4.2 构造 β 的置信域, 需要利用 plug-in 方法得到 β 极大经验似然估计协方差矩阵的相合估计. 由矩方法, Γ 和 Ω 的相合估计分别为

$$\hat{\Gamma} = \frac{1}{n}\sum_{i=1}^n \left\{\frac{\partial\eta_i^*\left(\Sigma_{uu}, \hat{\beta}; W_i, Y_i\right)}{\partial\beta}\right\},$$

$$\hat{\Omega} = \frac{1}{n}\sum_{i=1}^n \left\{\eta_i^*\left(\Sigma_{uu}, \hat{\beta}; W_i, Y_i\right)\eta_i^{*\mathrm{T}}\left(\Sigma_{uu}, \hat{\beta}; W_i, Y_i\right)\right\}.$$

然而, 在实际问题中 Σ_{uu} 通常是未知的. 当 Σ_{uu} 未知时, 测量误差协方差矩阵 Σ_{uu} 能够通过重复抽样进行估计 (Liang et al., 1999). 观测 $W_{ij} = X_i + U_{ij}, j = 1, \cdots, m_i$. 令 $\overline{W}_i$ 表示重复抽样的样本均值, 则 Σ_{uu} 一个相合的, 无偏的矩估计定义为

$$\hat{\Sigma}_{uu} = \frac{\sum_{i=1}^n\sum_{j=1}^{m_i}(W_{ij} - \overline{W}_i)(W_{ij} - \overline{W}_i)^{\mathrm{T}}}{\sum_{i=1}^n(m_i - 1)}.$$

注意到 $\mathrm{Cov}(\overline{W}_i)=m_i^{-1}\Sigma_{uu}$, 分别用 $\overline{W}_i$ 和 $m_i^{-1}\hat{\Sigma}_{uu}$ 替换 W_i 和 Σ_{uu}, 则得到无偏得分函数为 $\eta_i^*\left(m_i^{-1}\hat{\Sigma}_{uu},\beta;\overline{W}_i,Y_i\right)$, 相应的纠偏的经验对数似然比函数表示为 $l\left(\hat{\Sigma}_{uu},\beta\right)$. 当 $n\to\infty$ 时, 假设 $\dfrac{1}{n}\sum\limits_{i=1}^{n}m_i^{-1}$ 收敛到一个有界的常数.

定理 2.4.3 在定理 2.4.1 的条件下, 当 $n\to\infty$ 时, 仍有下面的结论

$$l\left(\hat{\Sigma}_{uu},\beta\right)\xrightarrow{L}\chi_p^2.$$

极大化 $\left\{-l\left(\hat{\Sigma}_{uu},\beta\right)\right\}$, 获得参数 β 的估计, 记为 $\tilde{\beta}$. 下面定理给了估计量 $\tilde{\beta}$ 的渐近分布.

定理 2.4.4 在定理 2.4.1 的条件下, 如果 Γ^* 是正定矩阵, 当 $n\to\infty$ 时, 则有

$$\sqrt{n}(\tilde{\beta}-\beta)\xrightarrow{L}N\left(0,\Gamma^{*-1}\Omega^*(\Gamma^{*-1})^{\mathrm{T}}\right),$$

其中

$$\Gamma^*=\lim_{n\to\infty}\frac{1}{n}\sum_{i=1}^{n}E\left\{\partial\eta^*\left(m_i^{-1}\Sigma_{uu},\beta;\overline{W}_i,Y_i\right)/\partial\beta\right\}$$

和

$$\Omega^*=\lim_{n\to\infty}\frac{1}{n}\sum_{i=1}^{n}E\left\{\eta^*\left(m_i^{-1}\Sigma_{uu},\beta;\overline{W}_i,Y_i\right)\eta^{*\mathrm{T}}\left(m_i^{-1}\Sigma_{uu},\beta;\overline{W}_i,Y_i\right)\right\}.$$

Γ^* 和 Ω^* 的相合估计容易定义为

$$\hat{\Gamma}^*=\frac{1}{n}\sum_{i=1}^{n}\left\{\frac{\partial\eta_i^*\left(\dfrac{1}{m_i}\hat{\Sigma}_{uu},\tilde{\beta};\overline{W}_i,Y_i\right)}{\partial\beta}\right\},$$

$$\hat{\Omega}^*=\frac{1}{n}\sum_{i=1}^{n}\left\{\eta_i^*\left(\frac{1}{m_i}\hat{\Sigma}_{uu},\tilde{\beta};\overline{W}_i,Y_i\right)\eta_i^{*\mathrm{T}}\left(\frac{1}{m_i}\hat{\Sigma}_{uu},\tilde{\beta};\overline{W}_i,Y_i\right)\right\}.$$

2.4.2 偏 profile 经验似然方法

当 β 是 2 维以上的向量时, 很难直接画出 β 的联合置信域. 然而, 可使用偏 profile 经验似然方法构造 β 某个分量的置信区间. 令 $e_{r,p}$ 表示第 r 个位置为 1, 其他位置为 0 的 p 维单位向量, $r=1,\cdots,p$. 类似于 Xue 和 Zhu (2007), 当 Σ_{uu} 是已知和未知时, β 的第 r 个分量 β_r 的估计分别表示为 $\hat{\beta}_r=e_{r,p}^{\mathrm{T}}\hat{\beta}$ 和 $\tilde{\beta}_r=e_{r,p}^{\mathrm{T}}\tilde{\beta}$. 记无偏得分函数分别为 $\eta_i^*(\beta)\equiv\eta_i^*\left(\Sigma_{uu},\beta;W_i,Y_i\right)$ 和 $\hat{\eta}_i^*(\beta)\equiv\eta_i^*\left(\dfrac{1}{m_i}\hat{\Sigma}_{uu},\beta;\overline{W}_i,Y_i\right)$. 令

$$\eta_{ir}^*(\beta_r)=e_{r,p}^{\mathrm{T}}\hat{\Gamma}^{-1}\eta_i^*(\hat{\beta}_1,\cdots,\hat{\beta}_{r-1},\beta_r,\hat{\beta}_{r+1},\cdots,\hat{\beta}_p),$$

$$\hat{\eta}_{ir}^*(\beta_r) = e_{r,p}^{\mathrm{T}} \hat{\Gamma}^{*-1} \hat{\eta}_i^*(\tilde{\beta}_1, \cdots, \tilde{\beta}_{r-1}, \beta_r, \tilde{\beta}_{r+1}, \cdots, \tilde{\beta}_p).$$

则 β_r 分别在矩阵 Σ_{uu} 已知和未知两种情况下的偏 profile 经验对数似然比函数为

$$l_r(\Sigma_{uu}, \beta_r) = -2 \max \left\{ \sum_{i=1}^n \log(np_i) \middle| p_i \geqslant 0, \sum_{i=1}^n p_i = 1, \sum_{i=1}^n p_i \eta_{ir}^*(\beta_r) = 0 \right\}$$

和

$$l_r\left(\hat{\Sigma}_{uu}, \beta_r\right) = -2 \max \left\{ \sum_{i=1}^n \log(np_i) \middle| p_i \geqslant 0, \sum_{i=1}^n p_i = 1, \sum_{i=1}^n p_i \hat{\eta}_{ir}^*(\beta_r) = 0 \right\}.$$

在定理 2.4.1 和定理 2.4.3 的条件下, 将证明 $l_r(\Sigma_{uu}, \beta_r)$ 和 $l_r\left(\hat{\Sigma}_{uu}, \beta_r\right)$ 的渐近分布, 即证明渐近收敛到标准 χ_1^2 分布.

定理 2.4.5 在定理 2.4.1 的条件下, 当 $n \to \infty$ 时, 有

$$l_r(\Sigma_{uu}, \beta_r) \xrightarrow{L} \chi_1^2.$$

在定理 2.4.3 的条件下, 有

$$l_r\left(\hat{\Sigma}_{uu}, \beta_r\right) \xrightarrow{L} \chi_1^2.$$

应用定理 2.4.5, 能够构造 β_r 的置信度为 $1-\alpha$ 的逼近置信区间. 此外, 定理 2.4.1~定理 2.4.5 的详细证明见 Yang 等 (2015).

注 2.4.2 上面的方法也可用于构造任意两个不同分量 (β_r, β_s) 的置信域. 方法如下: ①令 $e_{r,s,p}$ 表示第 r 和第 s 个位置为 1, 其他位置为 0 的 p 维向量. 这时用 $e_{r,s,p}$ 替换 $e_{r,p}$, 进而在得分函数 $\eta_i^*(\beta)$ 或 $\hat{\eta}_i^*(\beta)$ 中除了 β 的分量 (β_r, β_s) 外, 其他分量用极大经验似然估计替换, 得到 (β_r, β_s) 的得分函数 $\eta_{irs}^*(\beta_r, \beta_s)$ 或 $\hat{\eta}_{irs}^*(\beta_r, \beta_s)$; ② 证明 $l_{rs}(\Sigma_{uu}, \beta_r, \beta_s)$ 和 $l_{rs}(\Sigma_{uu}, \beta_r, \beta_s)$ 渐近收敛到标准 χ_2^2 分布, 然后利用理论结果构造 (β_r, β_s) 的经验似然置信域.

2.4.3 模拟研究

本节通过模拟研究和实际数据分析研究所提方法的有限样本性质, 模拟研究报告了泊松测量误差回归模型和 Logistic 测量误差回归模型的结果, 其他测量误差模型的模拟结果非常类似, 此处不再提供. 模拟研究中仅仅考虑了 β 的维数为 1 维和 2 维的情形. 当 β 是 3 维向量时, 即 $\beta = (\beta_1, \beta_2, \beta_3)^{\mathrm{T}}$, 能够使用定理 2.4.5 分别构造 β_1, β_2 和 β_3 的置信区间, 在实际数据分析中, 考虑了 3 维的情形. 在模拟研究中考虑了如下的问题.

(1) 比较纠偏经验对数似然比 (corrected empirical log-likelihood ratio, CELR) 和自然经验对数似然比 (naive empirical log-likelihood ratio, NELR) 的经验分布, 其

中自然经验对数似然比是忽略测量误差, 即直接用观测变量 W 代替变量 X 计算得到的, 并假定理论结果具有定理 2.4.3 的结论.

(2) 比较三种方法所构造的 β 的置信域: 忽略测量误差的自然的经验似然方法 (NEL), 基于定理 2.4.3 的纠偏的经验似然方法 (CEL) 和基于定理 2.4.4 的正态逼近方法 (NA).

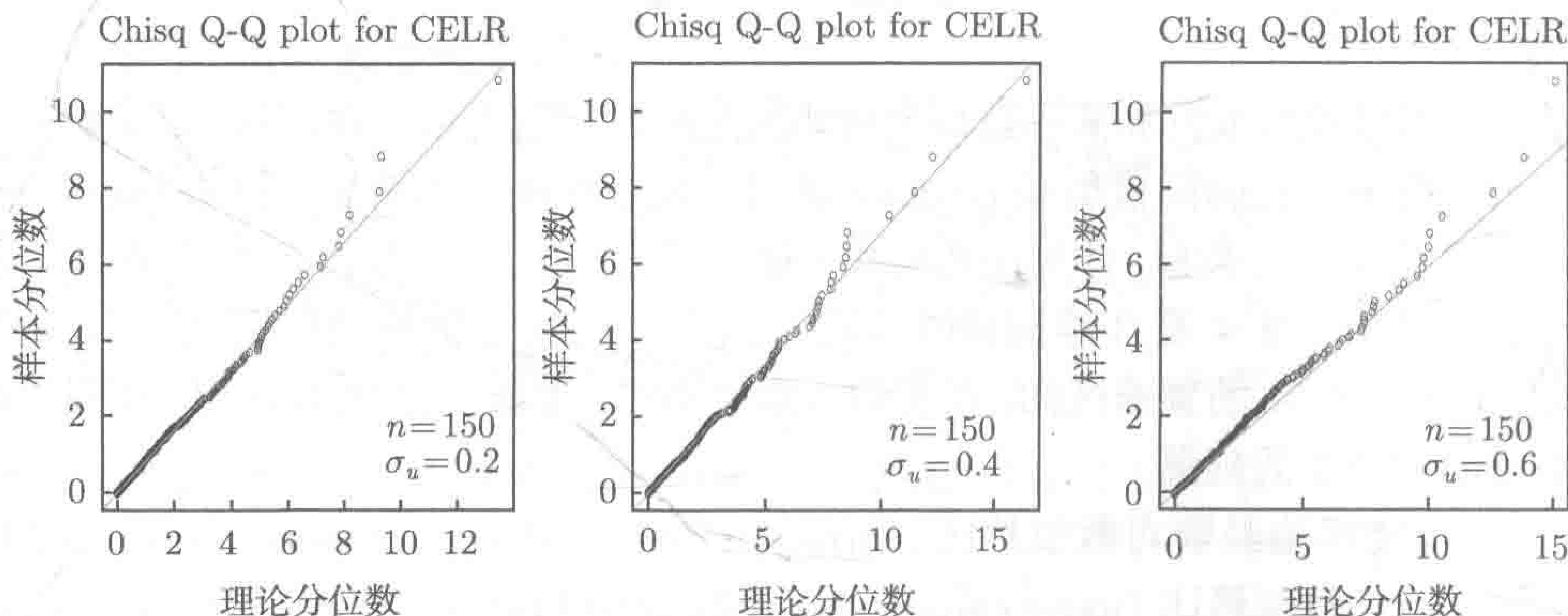
例 2.4.1 (一维数据情形)　考虑泊松测量误差回归模型, 响应变量 Y 来自于泊松分布 Poisson(μ), 且

$$\log(\mu) = X\beta, \quad W = X + U,$$

其中 $\beta = 1$, X 从 $N(1,1)$ 中产生, U 来自于 $N(0,\sigma_u^2)$. 取 $\sigma_u = 0.2, 0.4$ 和 0.6 三种情况, 表示不同的测量误差水平. σ_u 未知, 产生 W 的两次重复来估计 σ_u. 样本量大小取 $n = 100, 150$ 和 200, 对每种情况, 重复 2000 次模拟试验评价方法的完成情况.

首先在样本量大小 $n = 150$ 时, 比较了 $l(\beta)$ 渐近收敛到 χ_1^2 分布的有限样本性质. 图 2.4.1 给了 CELR 和 NELR 的 Q-Q 图. 明显可以发现, 在不同的测量误差水平下, CELR 的分布能够被 χ_1^2 分布逼近, 而 NELR 不能很好地被 χ_1^2 分布逼近, 这是因为它是基于一个有偏的得分函数.

为了评价 β 置信区间的完成情况, 模拟研究中计算了 95% 置信区间的平均长度和相应的覆盖概率. 模拟结果报告在表 2.4.1 中. 从表 2.4.1 中的结果可以看到: ① 基于 NEL 的置信区间是有偏的, 并且当测量误差水平变高时, 而覆盖概率变小; ②基于 CEL 和 NA 的置信区间的平均长度和覆盖概率依赖于测量误差和样本量大小. 对于固定的样本量大小, 当测量误差水平增大时, 置信区间的长度变长而覆盖概率变小. 对于固定的测量误差水平, 当样本量增大时, 置信区间的平均长度变短而覆盖概率变大; ③ 对比正态逼近方法, CEL 方法得到的置信区间有较小的区间长度和较高的覆盖概率.



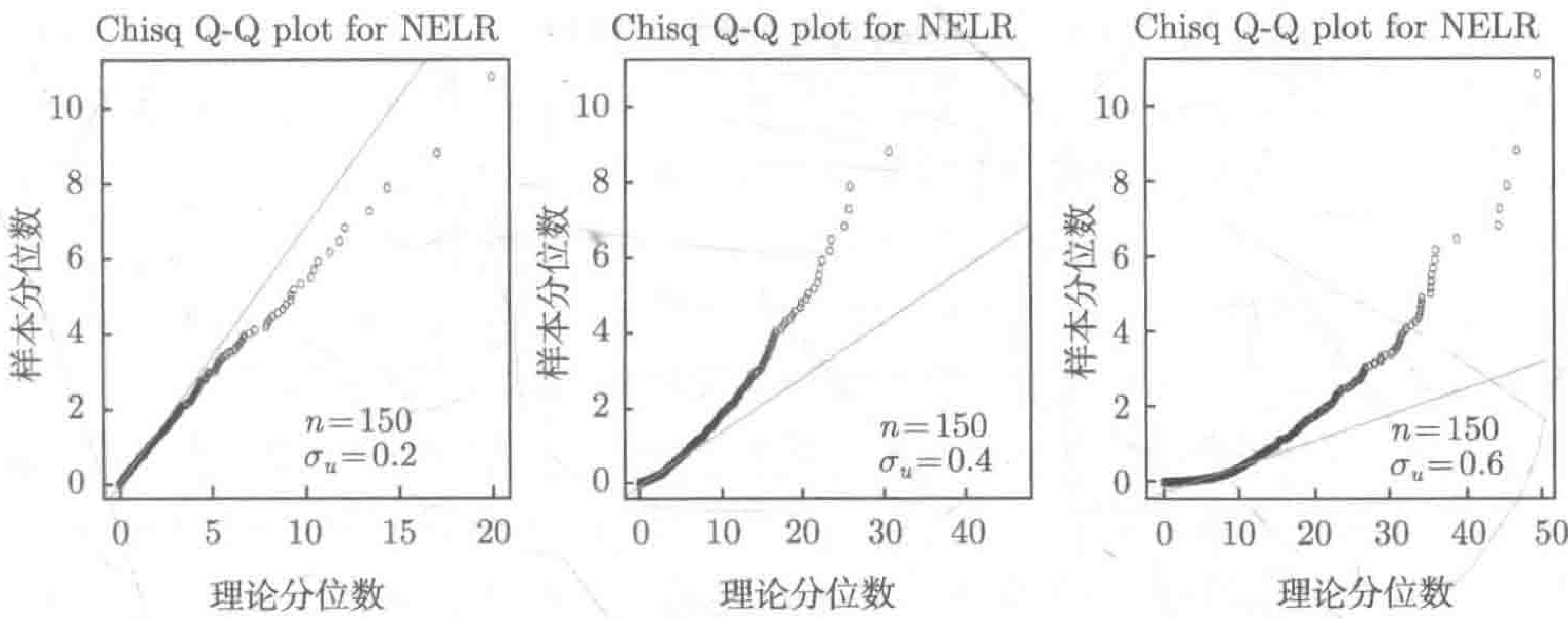


图 2.4.1 对于泊松测量误差回归模型, 当 $n = 150$, $\sigma_u = 0.2, 0.4$ 和 0.6 时, CELR 和 NELR 的 Q-Q 图

表 2.4.1 当置信水平为 0.95 时, 泊松测量误差回归模型中参数 β 基于 NEL, CEL 和 NA 三种方法置信区间的平均区间长度和覆盖概率

n	σ_u	平均区间长度			覆盖概率		
		NEL	CEL	NA	NEL	CEL	NA
100	0.2	0.0895	0.0907	0.0925	0.8966	0.9142	0.9048
	0.4	0.1102	0.1169	0.1211	0.7850	0.9108	0.8760
	0.6	0.1466	0.1537	0.1582	0.5610	0.8834	0.8506
150	0.2	0.074	0.0755	0.0774	0.9075	0.9305	0.9110
	0.4	0.0954	0.1013	0.1024	0.7370	0.9115	0.9030
	0.6	0.1284	0.1326	0.1342	0.4320	0.9040	0.8730
200	0.2	0.0661	0.0666	0.0669	0.8970	0.9310	0.9190
	0.4	0.0857	0.0897	0.0902	0.6890	0.9135	0.9040
	0.6	0.1203	0.1178	0.1202	0.3455	0.9065	0.8980

例 2.4.2 (二维数据情形) 数据从下面的 Logistic 测量误差回归模型中产生

$$\text{logit}(Y = 1|X) = X^{\mathrm{T}}\beta, \quad W = X + U,$$

其中 $\beta = (0.4, 0.8)^{\mathrm{T}}$, X_1 和 X_2 来自于 $N(0,1)$, U 来自于均值为 0, 协方差矩阵为 $\Sigma_{uu} = 0.4^2 I_2$ 的二维正态分布. Σ_{uu} 未知, 重复两次产生 W 来估计协方差矩阵 Σ_{uu}. 当样本量取 $n = 100, 150$ 和 200 时, 计算 CELR 和 NELR 的经验分布, 图 2.4.2 给出了经验分布的 Q-Q 图. 从图 2.4.2 可以看出, CELR 的分布比 NELR 的分布更好地逼近了 χ_2^2 分布.

图 2.4.2　对于 Logistic 测量误差回归模型, 当 $\sigma_u = 0.4$, $n = 100, 150$ 和 200 时, CELR 和 NELR 的 Q-Q 图

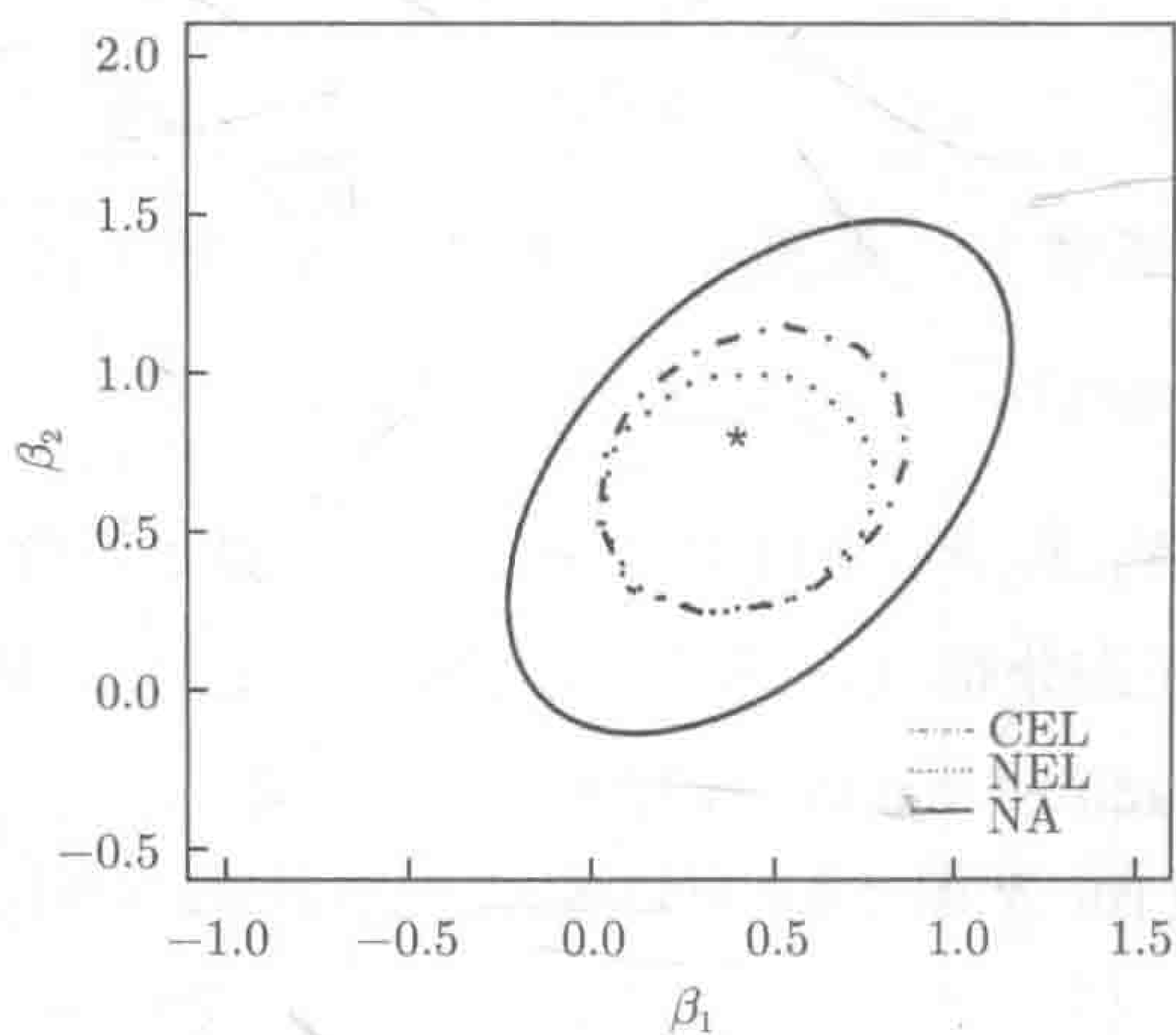


图 2.4.3　当 $n = 150$ 和 $\sigma_u = 0.4$ 时, (β_1, β_2) 的 95%的置信域, 其中虚线表示自然的经验似然方法 (NEL), 点断线表示纠偏的经验似然方法 (CEL), 实线表示正态逼近的方法 (NA)

当 $n = 150$ 和 $\sigma_u = 0.4$ 时, 图 2.4.3 计算了基于 2000 次重复模拟试验, β 的平均置信域和覆盖概率. 图 2.4.3 显示比较正态逼近方法, 纠偏的经验似然方法给了较小的置信域. 自然的经验似然方法 (NEL), 纠偏的经验似然方法 (CEL) 和正态逼近方法 (NA) 的经验覆盖概率分别为 0.902, 0.918 和 0.908.

例 2.4.3 (实际数据例子) 分析艾滋病临床试验数据 (ACTG 175, Huang and Wang, 2001) 来说明所提方法, 该数据集可从 R 程序包 “speff2trial” 中下载. 该随机试验的目的是评价具有一个或两个核苷艾滋病感染成年人的治疗情况, 主要对患者 200 到 500 之间 CD4 细胞进行筛选检查, 以及是否有艾滋病病史进行检查 (Hammer et al., 1996). 感兴趣的问题是研究正确基准 CD4 细胞数和静脉注射毒品史并有临床症状的 HIV 感染之间的关系. 分析 885 个抗逆转录病毒药物初治患者, 这些患者已经复制了基准的 CD4 细胞数的测量. 图 2.4.4 显示了 885 个患者复制的 CD4 细胞数的测量. 采用 Logistic 回归模型分析该数据集, 其中 Y 是响应变量, 表示有无临床症状的 HIV 感染, 其中 $Y = 0$ 表示无临床症状的 HIV 感染, $Y = 1$ 表示有临床症状的 HIV 感染. 在考虑的模型中有三个协变量, 其中 $X_1 = 1$ 表示截距项, X_2 表示正确基准 CD4 细胞数的对数, 即取 log(CD4), 而 CD4 细胞数的测量具有测量误差, 通过重复试验, 可以确定测量误差的方差为 $\hat{\sigma}_u^2 = 0.0565$. X_3 表示静脉注射毒品的历史 (其中, 0 表示没有静脉注射毒品的历史, 1 表示有静脉注射毒品的历史). 基于自然的经验似然方法 (NEL), 纠偏的经验似然方法 (CEL) 和正态逼近方法 (NA), 表 2.4.2 报告了 β 的估计和置信区间. 表 2.4.2 的结果可以看出, 静脉注射毒品的历史对是否有临床症状的 HIV 感染没有显著性的影响, 纠偏的经验似然方法有较小的置信区间. 当测量误差被考虑的时候, 惊奇地发现在 CD4 细胞数和是否有临床症状的 HIV 感染之间有某种较强的负相关.

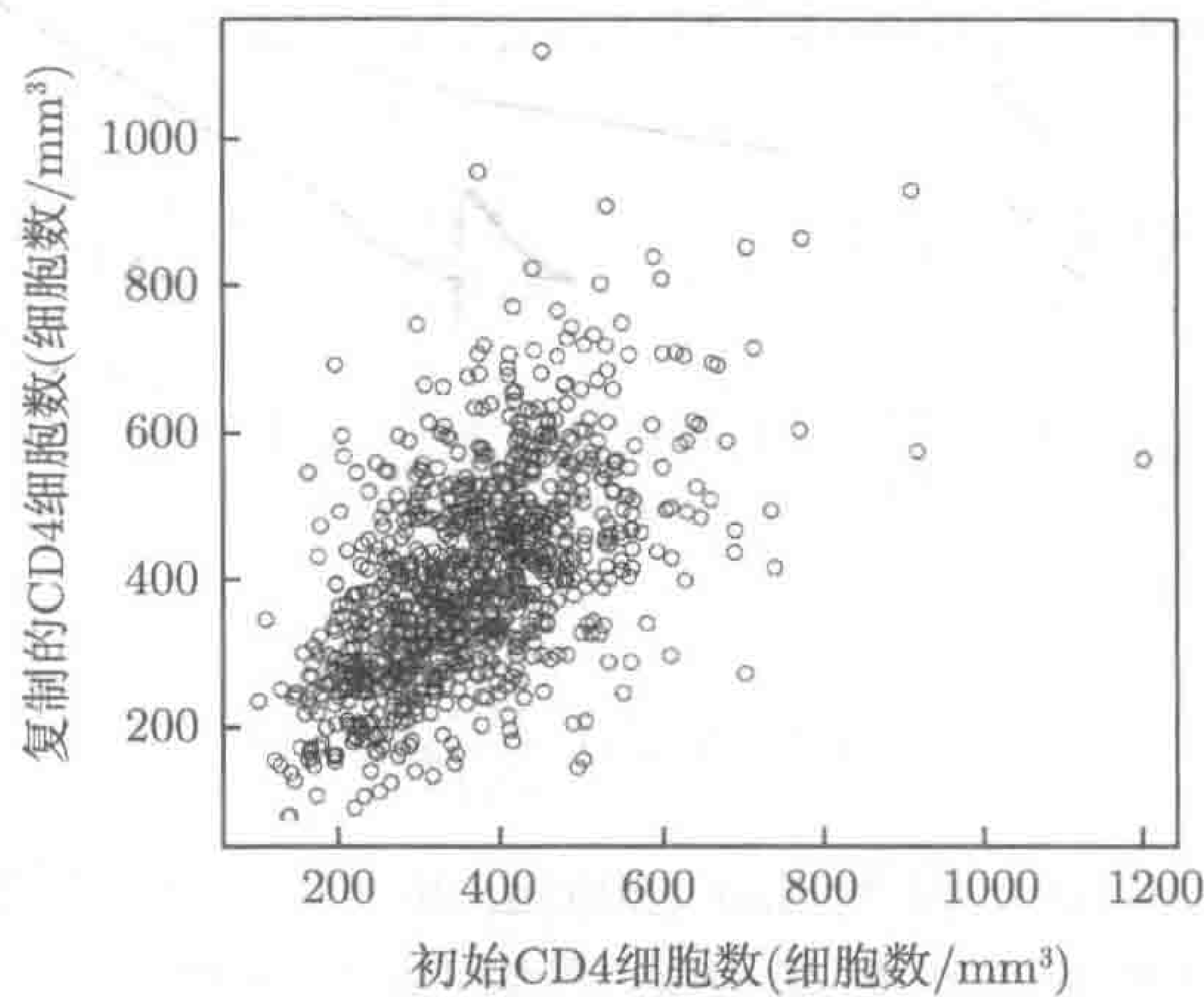


图 2.4.4 在 ACTG 175 研究中, 885 个抗逆转录病毒药物初治患者的复制的基准 CD4 细胞数的测量

表 2.4.2　在 ACTG 175 研究中, β 的估计和置信区间的结果

β	估计		置信区间		
	NEL	CEL/NA	NEL	CEL	NA
β_1	4.8929	8.1066	(4.1929,5.5929)	(7.5808,8.5074)	(7.6182,8.5949)
β_2	-1.1173	-1.6656	$(-1.6693, -0.4173)$	$(-1.8041, -1.5797)$	$(-2.1173, -1.2138)$
β_3	-0.0469	-0.0038	$(-0.6568, 0.5257)$	$(-0.6401, 0.5921)$	$(-0.6212, 0.6136)$

2.5　协变量调整回归模型的估计方法

在 2.1.3 节介绍了协变量调整线性回归模型 (2.1.7), 本节针对模型 (2.1.7), 介绍几种流行的估计方法, 如 binning 方法、最近邻 binning 方法、局部线性估计方法和直接估计方法.

2.5.1　binning 方法

Sentürk 和 Müller (2005a) 研究了模型 (2.1.7), 并提出了一个基于 binning 的估计方法, Sentürk 和 Müller (2006) 证明了估计量的渐近性质. 下面详细介绍 Sentürk 和 Müller (2005a) 所提出的 binning 方法.

假设 $\{(\widetilde{Y}_i, \widetilde{X}_i, U_i), i = 1, \cdots, n\}$ 是来自于模型 (2.1.7) 的可观测样本, 其中 $\widetilde{X}_i = (\widetilde{X}_{i1}, \cdots, \widetilde{X}_{ip})^{\mathrm{T}}$ 是 p 维可观测的协变量. 由模型的假设条件: $E(\varepsilon) = 0$, 以及对于 $r = 1, \cdots, p$, (X_r, U, ε) 是相互独立的, 且 (Y, U) 也是独立的. 在给定观测协变量 $\widetilde{X}_i$ 和混淆变量 U_i 的条件下, 计算可观测的响应变量 $\widetilde{Y}_i$ 的条件期望为

$$
\begin{aligned}
E(\widetilde{Y}_i|\widetilde{X}_i, U_i) &= E\{\phi(U_i)Y_i|\psi_1(U_i)X_{i1}, \cdots, \psi_p(U_i)X_{ip}, U_i\} \\
&= \phi(U_i)E\left\{\beta_0 + \sum_{r=1}^{p} X_{ir}\beta_r + \varepsilon_i \middle| \psi_1(U_i)X_{i1}, \cdots, \psi_p(U_i)X_{ip}, U_i\right\} \\
&= \phi(U_i)\beta_0 + \phi(U_i)\sum_{r=1}^{p}\beta_r\frac{\psi_r(U_i)X_{ir}}{\psi_r(U_i)} \\
&=: \alpha_0(U_i) + \sum_{r=1}^{p}\alpha_r(U_i)\widetilde{X}_{ir},
\end{aligned}
$$

其中记

$$
\alpha_0(U_i) = \phi(U_i)\beta_0, \quad \alpha_r(U_i) = \beta_r\frac{\phi(U_i)}{\psi_r(U_i)}, \quad r = 1, \cdots, p. \tag{2.5.1}
$$

由上面的推导, Sentürk 和 Müller (2005a, 2006) 利用可观测的响应变量 $\widetilde{Y}_i$ 和协变量 $\{\widetilde{X}_i, U_i\}$ 把模型 (2.1.7) 转换成如下的变系数模型:

$$
\widetilde{Y}_i = \alpha_0(U_i) + \sum_{r=1}^{p}\alpha_r(U_i)\widetilde{X}_{ir} + \widetilde{\varepsilon}_i, \tag{2.5.2}
$$

其中 $\widetilde{\varepsilon}_i = \phi(U_i)\varepsilon_i$. 对于协变量调整线性回归模型 (2.1.7), 主要的目的是利用可观测数据得到模型中的未知参数 $\beta_0, \beta_1, \cdots, \beta_p$ 的相合估计, 并证明估计量的渐近性质. Sentürk 和 Müller (2005a, 2006) 把模型 (2.1.7) 转换成变系数模型 (2.5.2), 这时文献中有很多统计方法能够很好地估计模型中的系数函数 $\alpha_0(\cdot), \alpha_1(\cdot), \cdots, \alpha_p(\cdot)$, Sentürk 和 Müller(2005a, 2006) 采用 Fan 和 Zhang (2000) 研究纵向数据变系数模型时提出的 binning 方法. 由失真函数的约束条件 $E\{\phi(U)\} = 1, E\{\psi_r(U)\} = 1, r = 1, \cdots, p$, 并利用式 (2.5.1) 所定义的结构, 一旦利用 binning 方法得到系数函数 $\alpha_0(\cdot), \alpha_1(\cdot), \cdots, \alpha_p(\cdot)$ 的估计后, 进行加权平均即可得到参数 $\beta_0, \beta_1, \cdots, \beta_p$ 的估计. 具体程序如下:

步骤 1 假设变量 U 的支撑集有界, 即 $-\infty < a \leqslant U \leqslant b < \infty$, 其中实数 $a < b$. 将区间 $[a, b]$ 等分成 m 个区间, 记为 $B_1, \cdots, B_m$. 给定 m, 每个区间 $B_j (j = 1, \cdots, m)$ 是固定的, 但是落在每个区间 B_j 中的变量 U_i 的个数是随机的. 令 L_j 表示落在第 j 个区间 B_j 中样本的个数, $j = 1, \cdots, m$, 并令第 j 个区间 B_j 中的数据集为

$$\begin{aligned} &\{(\widetilde{Y}^*_{jk}, \widetilde{X}^*_{jkr}, U^*_{jk}), k = 1, \cdots, L_j, r = 1, \cdots, p\} \\ =&\{(\widetilde{Y}_i, \widetilde{X}_{ir}, U_i), i = 1, \cdots, n, r = 1, \cdots, p, U_i \in B_j\}. \end{aligned}$$

为了方便表示估计量, 引进一些记号. 令 $U^*_j = (U^*_{j1}, \cdots, U^*_{jL_j})^{\mathrm{T}}$, $\widetilde{Y}^*_j = (\widetilde{Y}^*_{j1}, \cdots, \widetilde{Y}^*_{jL_j})^{\mathrm{T}}$, $\widetilde{X}^*_{jk} = (1, \widetilde{X}^*_{jk1}, \cdots, \widetilde{X}^*_{jkp})^{\mathrm{T}}, k = 1, \cdots, L_j$, 则定义 $\widetilde{X}^*_j = (\widetilde{X}^*_{j1}, \cdots, \widetilde{X}^*_{jL_j})^{\mathrm{T}}$ 为第 j 个区间 B_j 上 $L_j \times (p+1)$ 的可观测协变量矩阵. 这时, 基于第 j 个区间 B_j 上的数据集, 利用最小二乘方法, 可得

$$\hat{\alpha}_j = (\hat{\alpha}_{j0}, \hat{\alpha}_{j1}, \cdots, \hat{\alpha}_{jp})^{\mathrm{T}} = \left(\widetilde{X}^{*\mathrm{T}}_j \widetilde{X}^*_j\right)^{-1} \widetilde{X}^{*\mathrm{T}}_j \widetilde{Y}^*_j. \tag{2.5.3}$$

步骤 2 由式 (2.5.3) 得到区间 B_j 上的估计 $\hat{\alpha}_j$ 后, 然后进行加权平均, 进而得到参数 $\beta_0, \beta_1, \cdots, \beta_p$ 的估计. 回顾 L_j 表示样本 $U_1, \cdots, U_n$ 落入第 j 个区间 B_j 内样本的个数, 则有

$$\hat{\beta}_0 = \sum_{j=1}^{m} \frac{L_j}{n} \hat{\alpha}_{j0}, \quad \hat{\beta}_r = \frac{1}{\bar{\widetilde{X}}_r} \sum_{j=1}^{m} \frac{L_j}{n} \hat{\alpha}_{jr} \bar{\widetilde{X}}^*_{jr}, \quad r = 1, \cdots, p, \tag{2.5.4}$$

其中 $\bar{\widetilde{X}}^*_{jr} = \dfrac{1}{L_j} \sum\limits_{k=1}^{L_j} \widetilde{X}^*_{jkr}$, $\bar{\widetilde{X}}_r = \dfrac{1}{n} \sum\limits_{i=1}^{n} \widetilde{X}_{ir}, j = 1, \cdots, m, r = 1, \cdots, p$, 且 $\bar{\widetilde{X}}^*_{jr}$ 表示落入区间 B_j 的样本均值.

Sentürk 和 Müller (2006) 证明由式 (2.5.4) 定义的估计具有渐近正态性, 当 $n \to \infty$ 时, $m \to \infty, n/(m \log n) \to \infty$ 和 $m/\sqrt{n} \to \infty$, 以及一些正则条件下, 则有

$$\sqrt{n}(\hat{\beta}_r - \beta_r) \xrightarrow{L} N(0, \sigma^2_r), \quad 0 \leqslant r \leqslant p, \tag{2.5.5}$$

其中

$$
\begin{aligned}
\sigma_0^2 =& \sigma^2(\Lambda^{-1})_{(1,1)}E\{\phi^2(U)\} + \beta_0^2\text{Var}(\phi(U)),\\
\sigma_r^2 =& \frac{\beta_r^2[E(X_r^2)E\{\phi^2(U)\} - \{E(X_r)\}^2] + \sigma^2\{E(X_r)\}^2E\{\phi^2(U)\}(\Lambda^{-1})_{(r,r)}}{\{E(X_r)\}^2}\\
&-\frac{2\beta_r^2[E(X_r^2)E\{\phi(U)\psi_r(U)\} - \{E(X_r)\}^2] + \beta_r^2\text{Var}(\widetilde{X}_r)}{\{E(X_r)\}^2}, \quad r = 1, \cdots, p,
\end{aligned}
$$

$$
\Lambda = \begin{pmatrix} 1 & E(X_1) & \cdots & E(X_p) \\ E(X_1) & E(X_1^2) & \cdots & E(X_1X_p) \\ \vdots & \vdots & & \vdots \\ E(X_p) & E(X_1X_p) & \cdots & E(X_p^2) \end{pmatrix}
$$

为非奇异的 $(p+1)\times(p+1)$ 的矩阵. $A_{(i,j)}$ 表示矩阵 A 的第 (i,j) 个位置上的元素.

对于 binning 方法, 至关重要的一个问题是如何确定区间个数 m 的大小. Sentürk 和 Müller (2006) 建议在每个区间 B_j 中至少包含 $p+1$ 个样本点, 保证在每个区间中有足够的数据能够得到最小二乘估计 $\hat{\alpha}_j$. 如果某个区间中的样本点少于 $p+1$ 个, 这时可将该区间与邻近的区间进行合并. 例如, 对于样本量为 508 的数据集, 若每个区间的样本点数选取为 14 个时, 可以保证该区间能够得到最小二乘估计 $\hat{\alpha}_j$, 则区间个数可以取为 $m = \lfloor 508/14 \rfloor = 36$, 其中 $\lfloor\cdot\rfloor$ 表示向下取整.

2.5.2 最近邻 binning 方法

在 2.5.1 节介绍的 binning 方法的步骤 1 中, 对变量 U 的支撑集 $[a,b]$ 的划分, 是等分成 m 个区间, 记为 $B_1, \cdots, B_m$. 给定 m 时, 每个区间 $B_j (j=1,\cdots,m)$ 是固定的, 但是落在每个区间 B_j 中的变量 U_i 的个数是随机的. Sentürk 和 Nguyen (2006) 指出一旦每个区间 B_j 固定以后, 很难保证每个区间中包含的样本点数不少于 $p+1$ 个, 这时对样本点少于 $p+1$ 个的区间, 很难得到很好的拟合效果. 为了解决这个问题, Sentürk 和 Nguyen (2006) 提出了最近邻 binning 方法. 区别于 binning 方法, 最近邻 binning 方法在划分支撑集 $[a,b]$ 时, 要求每个区间的长度和边界点是随机的, 但是每个区间包含的观测样本点个数 L 相同, 并要求 $L \geqslant p+1$.

具体过程为: 利用最近邻的思想, 首先把混淆变量 $U_i, i=1,\cdots,n$, 从小到大进行排序, 然后固定每个区间包含的样本量大小 $L \geqslant p+1$, 把变量 U 的支撑集 $[U_{(1)}, U_{(n)}]$ 划分成 $m = n/L$ 个区间, 其中 $U_{(1)} = \min\{U_1,\cdots,U_n\}$ 和 $U_{(n)} = \max\{U_1,\cdots,U_n\}$. 一旦把区间 B_j $(j=1,\cdots,m)$ 划分好后, 最近邻 binning 估计方法的估计步骤就相同于 binning 方法的估计步骤, 最后可以得到未知参数 $\beta_0, \beta_1, \cdots, \beta_p$ 的估计.

2.5.3 局部线性估计方法

2.5.1 节 binning 估计步骤 1 中关于系数函数 $\alpha_0(\cdot),\alpha_1(\cdot),\cdots,\alpha_p(\cdot)$ 的估计, 可以采用非参数的一些估计方法进行估计, 如核光滑方法、局部多项式估计方法、样条估计方法和小波估计方法等. Sentürk 和 Nguyen (2006) 把 Fan 和 Gijbels (1996) 提出的局部线性估计方法用于步骤 1 中的系数函数的估计, 提出了局部线性的估计方法.

步骤 1 首先将未知系数函数 $\alpha_r(U)$ 在 u 的邻域内进行如下的局部线性逼近:

$$\alpha_r(U)\approx\alpha_r(u)+\alpha_r'(u)(U-u)=:a_r+b_r(U-u),\quad r=0,1,\cdots,p, \tag{2.5.6}$$

其中 $\alpha_r'(\cdot)$ 表示 $\alpha_r(\cdot)$ 的一阶导数. 由式 (2.5.6), 定义局部加权最小二乘目标函数如下:

$$\sum_{i=1}^{n}\left[\widetilde{Y}_i-\sum_{r=0}^{p}\{a_r+b_r(U_i-u)\}\widetilde{X}_{ir}\right]^2 K_h(U_i-u), \tag{2.5.7}$$

其中 $\widetilde{X}_{i0}\equiv 1$, $K_h(\cdot)=K(\cdot/h)/h$, 且 h 是窗宽, $K(\cdot)$ 是选定的核函数. 令 $e_{2r+1,2(p+1)}$ 表示长度 $2(p+1)$, 且第 $2r+1$ 元素为 1, 而其余元素全为 0 的单位向量, $\widetilde{\mathcal{Y}}=(\widetilde{Y}_1,\cdots,\widetilde{Y}_n)^{\mathrm{T}}$, $\mathcal{W}=\mathrm{diag}\{K_h(U_1-u),\cdots,K_h(U_n-u)\}$ 是一个 $n\times n$ 的对角矩阵, 并且

$$\widetilde{\mathcal{X}}=\begin{pmatrix} 1 & U_1-u & \widetilde{X}_{11} & \widetilde{X}_{11}(U_1-u) & \cdots & \widetilde{X}_{1p} & \widetilde{X}_{1p}(U_1-u)\\ \vdots & \vdots & \vdots & \vdots & & \vdots & \vdots\\ 1 & U_n-u & \widetilde{X}_{n1} & \widetilde{X}_{n1}(U_n-u) & \cdots & \widetilde{X}_{np} & \widetilde{X}_{np}(U_n-u)\end{pmatrix}.$$

极小化局部加权最小二乘目标函数 (2.5.7), 可以得到 $\alpha_r(u)$ 的估计如下:

$$\hat{\alpha}_r(u)=\hat{a}_r=e_{2r+1,2(p+1)}^{\mathrm{T}}\left(\widetilde{\mathcal{X}}^{\mathrm{T}}\mathcal{W}\widetilde{\mathcal{X}}\right)^{-1}\widetilde{\mathcal{X}}^{\mathrm{T}}\mathcal{W}\widetilde{\mathcal{Y}},\quad r=0,1,\cdots,p. \tag{2.5.8}$$

步骤 2 最终的目的是估计未知参数 $\beta_0,\beta_1,\cdots,\beta_p$, 类似于 2.5.1 节的步骤 2, 只需要对步骤 1 中得到的系数函数的估计作平均即可获得未知参数的估计. 为了更好地定义平均值, 记

$$\hat{\alpha}_r(U_i)=e_{2r+1,2(p+1)}^{\mathrm{T}}\left(\widetilde{\mathcal{X}}_i^{\mathrm{T}}\mathcal{W}_i\widetilde{\mathcal{X}}_i\right)^{-1}\widetilde{\mathcal{X}}_i^{\mathrm{T}}\mathcal{W}_i\widetilde{\mathcal{Y}},\quad r=0,1,\cdots,p,$$

其中 $\widetilde{\mathcal{X}}_i$ 和 $\mathcal{W}_i$ 是在 $\widetilde{\mathcal{X}}$ 和 $\mathcal{W}$ 中令 $u=U_i$. 这时定义回归参数 $\beta_0,\beta_1,\cdots,\beta_p$ 的估计为

$$\hat{\beta}_0=\frac{1}{n}\sum_{i=1}^{n}\hat{\alpha}_0(U_i),\quad \hat{\beta}_r=\frac{1}{\bar{\widetilde{X}}_r}\sum_{i=1}^{n}\frac{1}{n}\hat{\alpha}_r(U_i)\widetilde{X}_{ir},\quad r=0,1,\cdots,p. \tag{2.5.9}$$

Sentürk 和 Nguyen (2006) 证明了由式 (2.5.9) 定义的估计量的相合性, 即

$$\hat{\beta}_r - \beta_r = O_P(n^{-1/2}) + O_P(h^2), \quad r = 0, 1, \cdots, p.$$

局部线性估计方法最关键的问题是需要选取窗宽 h, Sentürk 和 Nguyen (2006) 提出采用广义交错核实 (generalized cross-validation, GCV) 准则进行选取, 关于 GCV 方法的更多讨论见 Wahba (1977), Craven 和 Wahba (1979). 由式 (2.5.8) 定义的估计 $\hat{\alpha}_j(u)$ 关于可观测响应变量 $\{\widetilde{Y}_i, i = 1, \cdots, n\}$ 是线性的, 则响应变量的第 i 个拟合值能够表达为响应变量的线性组合, 即

$$\begin{aligned}\hat{\widetilde{Y}}_i &= \sum_{j=1}^{n} \left\{ v_j^{(1)}(U_i) + v_j^{(3)}(U_i)\widetilde{X}_{i1} + \cdots + v_j^{(2p+1)}(U_i)\widetilde{X}_{ip} \right\} \widetilde{Y}_j \\ &=: \sum_{j=1}^{n} \left\{ v_j^*(U_i) \right\} \widetilde{Y}_j,\end{aligned}$$

其中, 对于固定的窗宽 h, 记

$$v^{(2r+1)}(U_i) = \left[v_1^{(2r+1)}(U_i), \cdots, v_n^{(2r+1)}(U_i) \right] = e_{2r+1,2(p+1)}^{\mathrm{T}} \left(\widetilde{\mathcal{X}}_i^{\mathrm{T}} \mathcal{W}_i \widetilde{\mathcal{X}}_i \right)^{-1} \widetilde{\mathcal{X}}_i^{\mathrm{T}} \mathcal{W}_i.$$

令 $\hat{\widetilde{\mathcal{Y}}} = \mathcal{V}\widetilde{\mathcal{Y}}$, 其中 $\mathcal{V}$ 的第 i 行为 $[v_1^*(U_i), v_2^*(U_i), \cdots, v_n^*(U_i)], i = 1, \cdots, n$. 定义残差平方和 $\mathrm{RSS}(h) = \|\hat{\widetilde{\mathcal{Y}}} - \widetilde{\mathcal{Y}}\|^2$, 则极小化下面的 GCV 目标函数, 可选取窗宽 h, 即

$$h_{\mathrm{gcv}} = \arg\min_{h} \frac{n^{-1}\mathrm{RSS}(h)}{[1 - n^{-1}\mathrm{tr}(\mathcal{V})]^2}.$$

Sentürk 和 Nguyen (2006) 在模拟研究中显示, 通过极小化 GCV 准则选取的窗宽 h_{gcv} 表现是稳定的, 并且局部线性估计方法得到的未知参数的估计量要比 binning 方法和最近邻 binning 方法的估计量有较小的均方误差, 特别是对小样本的情形.

2.5.4 直接估计方法

对模型 (2.1.7), Cui (2008) 提出了一种新的直接估计方法, 主要想法是: ① 基于可识别条件 (2.1.8), 利用观测数据 $\{(\widetilde{Y}_i, \widetilde{X}_i, U_i), i = 1, \cdots, n\}$, 分别得到未知失真函数 $\phi(u)$ 和 $\psi_r(u)$ 的相合估计, $r = 1, \cdots, p$; ② 通过关系式 $Y = \widetilde{Y}/\phi(U)$ 和 $X_r = \widetilde{X}_r/\psi_r(U)$, 分别将不可观测的变量 Y_i 和 $X_i = (X_{i1}, \cdots, X_{ip})^{\mathrm{T}}$ 估计出来, 得到恢复数据集, 记为 $\{(\hat{Y}_i, \hat{X}_i), i = 1, \cdots, n\}$; ③ 基于恢复数据集, 直接使用最小二乘方法, 便可得到未知参数 $\beta_0, \beta_1, \cdots, \beta_p$ 的相合估计.

直接估计方法的一个明显优点是, 当考虑非线性模型或更复杂的半参数模型时, 如简单的非线性模型 $Y = g(X_1, \cdots, X_p, \beta) + \varepsilon$, 直接估计方法仍然可行并是

非常有效的, 见 Cui 等 (2009). 因为直接估计方法是用恢复数据集 $\{(\hat{Y}_i, \hat{X}_{ir}), i = 1, \cdots, n, r = 1, \cdots, p\}$ 替代不可观测的数据 $\{(Y_i, X_{ir}), i = 1, \cdots, n, r = 1, \cdots, p\}$, 然后利用传统的非线性最小二乘方法获得模型中的未知参数的估计. 明显直接估计方法不依赖于模型的线性结构, 另外既然恢复数据是不可观测数据的相合估计, 在一定正则条件下, 基于恢复数据得到的参数估计也是相合的.

下面介绍由 Cui (2008), Cui 等 (2009) 提出的直接估计方法. 由失真函数约束条件 (2.1.8), 以及假设条件 (X_r, U, ε) 相互独立和 (Y, U) 相互独立, $r = 1, \cdots, p$, 则有

$$
\begin{aligned}
&E(\widetilde{Y}|U) = \phi(U)E(Y), \quad E(\widetilde{X}_r|U) = \psi_r(U)E(X_r), \\
&E(Y) = E(\widetilde{Y}), \quad E(X_r) = E(\widetilde{X}_r), \quad r = 1, \cdots, p.
\end{aligned}
$$

基于上述公式, 有

$$
\phi(U) = \frac{E(\widetilde{Y}|U)}{E(Y)}, \quad \psi_r(U) = \frac{E(\widetilde{X}_r|U)}{E(X_r)}, \quad r = 1, \cdots, p. \tag{2.5.10}
$$

为了方便, 令 $f(u)$ 表示 U 的密度函数, 并定义

$$
g_Y(U) = E(\widetilde{Y}|U)f(U), \quad g_r(U) = E(\widetilde{X}_r|U)f(U), \quad r = 1, \cdots, p.
$$

利用 Cui (2008) 与 Cui 等 (2009) 的直接估计思想, 使用观测数据 $\{(\widetilde{Y}_i, \widetilde{X}_i, U_i), 1 \leqslant i \leqslant n\}$ 和约束条件 (2.5.10) 能够估计失真函数 $\phi(U)$ 和 $\psi_r(U) (1 \leqslant r \leqslant p)$ 如下

$$
\begin{aligned}
\hat{\phi}(u) &= \frac{\frac{1}{n}\sum_{i=1}^{n} K_h(u - U_i)\widetilde{Y}_i}{\frac{1}{n}\sum_{i=1}^{n} K_h(u - U_i)} \times \frac{1}{\bar{\widetilde{Y}}} =: \frac{\hat{g}_Y(u)}{\hat{f}(u)} \times \frac{1}{\bar{\widetilde{Y}}}, \\
\hat{\psi}_r(u) &= \frac{\frac{1}{n}\sum_{i=1}^{n} K_h(u - U_i)\widetilde{X}_{ir}}{\frac{1}{n}\sum_{i=1}^{n} K_h(u - U_i)} \times \frac{1}{\bar{\widetilde{X}}_r} =: \frac{\hat{g}_r(u)}{\hat{f}(u)} \times \frac{1}{\bar{\widetilde{X}}_r}, \quad r = 1, \cdots, p,
\end{aligned} \tag{2.5.11}
$$

其中$K_h(\cdot) = K(\cdot/h)/h$, $K(\cdot)$ 是核函数, h 是窗宽, 且 $\bar{\widetilde{Y}} = \frac{1}{n}\sum_{i=1}^{n}\widetilde{Y}_i$, $\bar{\widetilde{X}}_r = \frac{1}{n}\sum_{i=1}^{n}\widetilde{X}_{ir}$.

在模型 (2.1.7) 中第二和第三个式子中, 用相合估计 $\hat{\phi}(\cdot)$ 和 $\hat{\psi}_r(\cdot)$ 分别代替未知的 $\phi(\cdot)$ 和 $\psi_r(\cdot)$, 这时可获得恢复数据集 $\{(\hat{Y}_i, \hat{X}_i), 1 \leqslant i \leqslant n\}$, 其中 $\hat{Y}_i = \widetilde{Y}_i/\hat{\phi}(U_i)$, $\hat{X}_{ir} = \widetilde{X}_{ir}/\hat{\psi}_r(U_i)$ 和 $\hat{X}_i = (\hat{X}_{i1}, \cdots, \hat{X}_{ip})^{\mathrm{T}}$. 利用直接估计方法获取恢复数据时, 窗

宽 h 的选取可以采用 Silverman (1986) 提出的拇指准则, 取为 $\hat{h} = \hat{\sigma}_U n^{-1/3}$, 其中 $\hat{\sigma}_U$ 表示样本 $U_1, \cdots, U_n$ 的样本标准差.

基于恢复数据集 $\{(\hat{Y}_i, \hat{X}_i), 1 \leqslant i \leqslant n\}$, 构造下面的最小二乘目标函数

$$Q(\beta_0, \beta_1, \cdots, \beta_p) = \sum_{i=1}^{n} \left(\hat{Y}_i - \beta_0 - \sum_{r=1}^{p} \hat{X}_{ir} \beta_r \right)^2. \tag{2.5.12}$$

极小化由式 (2.5.12) 所定义的最小二乘目标函数 $Q(\beta_0, \beta_1, \cdots, \beta_p)$, 可得未知参数 $\beta_0, \beta_1, \cdots, \beta_p$ 的最小二乘估计为

$$\begin{aligned} \begin{pmatrix} \hat{\beta}_0 \\ \hat{\beta}_1 \\ \vdots \\ \hat{\beta}_p \end{pmatrix} &= \left[\frac{1}{n} \sum_{i=1}^{n} \begin{pmatrix} 1 \\ \hat{X}_i \end{pmatrix}^{\otimes 2} \right]^{-1} \left[\frac{1}{n} \sum_{i=1}^{n} \begin{pmatrix} 1 \\ \hat{X}_i \end{pmatrix} \hat{Y}_i \right] \\ &= (\hat{\mathcal{X}}^{\mathrm{T}} \hat{\mathcal{X}})^{-1} \hat{\mathcal{X}}^{\mathrm{T}} \hat{\mathcal{Y}}, \end{aligned} \tag{2.5.13}$$

其中

$$\hat{\mathcal{X}} = \begin{pmatrix} 1 & \hat{X}_{11} & \hat{X}_{12} & \dots & \hat{X}_{1p} \\ 1 & \hat{X}_{21} & \hat{X}_{22} & \dots & \hat{X}_{2p} \\ \vdots & \vdots & \vdots & & \vdots \\ 1 & \hat{X}_{n1} & \hat{X}_{n2} & \dots & \hat{X}_{np} \end{pmatrix}, \quad \hat{\mathcal{Y}} = (\hat{Y}_1, \hat{Y}_2, \cdots, \hat{Y}_n)^{\mathrm{T}}.$$

Cui (2008) 在一定的正则条件下, 证明了由式 (2.5.13) 定义的最小二乘估计的渐近正态性, 即有

$$\sqrt{n}(\hat{\beta}_r - \beta_r) \xrightarrow{L} N(0, e_{r+1,p+1}^{\mathrm{T}} \Sigma e_{r+1,p+1}), \quad r = 0, 1, \cdots, p, \tag{2.5.14}$$

其中 $e_{l,p+1}$ 表示第 l 个元素为 1, 其余元素全为 0, 长度为 $p+1$ 的单位列向量, $\Sigma = \sigma^2 \Lambda^{-1} + \Lambda^{-1} \Xi \Lambda^{-1}$, 且

$$\begin{aligned} \Xi = &\sum_{l=1}^{p} \sum_{s=1}^{p} \frac{E(X_l X_s)}{E(X_l) E(X_s)} \pi_{X_l} \pi_{X_s}^{\mathrm{T}} \mathrm{Cov}(\psi_l(U), \psi_s(U)) \beta_l \beta_s \\ &- \sum_{l=1}^{p} \frac{E(X_l Y)}{E(X_l) E(Y)} \left[\pi_{X_l} \pi_Y^{\mathrm{T}} + \pi_Y \pi_{X_l}^{\mathrm{T}} \right] \mathrm{Cov}(\psi_l(U), \phi(U)) \beta_l \\ &+ \frac{E(Y^2)}{[E(Y)]^2} \pi_Y^{\otimes 2} \mathrm{Var}(\phi(U)), \end{aligned}$$

以及 $\pi_{X_l} = E\left[\left(1, X^{\mathrm{T}}\right)^{\mathrm{T}} X_l \right]$, $\pi_Y = E\left[\left(1, X^{\mathrm{T}}\right)^{\mathrm{T}} Y \right]$.

2.6 置信区间

2.6.1 渐近正态近似方法

通常情况下, 对于 $r=0,1,\cdots,p$, 可以利用 β_r 估计量的渐近正态分布 $\sqrt{n}(\hat{\beta}_r-\beta_r)\xrightarrow{L}N(0,\sigma^2_{\beta_r})$ 来构造未知参数 β_r 的置信水平为 $1-\alpha$ 的置信区间如下

$$\left[\hat{\beta}_r-\frac{\hat{\sigma}_{\beta_r}}{\sqrt{n}}z_{1-\alpha/2},\hat{\beta}_r+\frac{\hat{\sigma}_{\beta_r}}{\sqrt{n}}z_{1-\alpha/2}\right], \tag{2.6.1}$$

其中 $\hat{\sigma}^2_{\beta_r}$ 是渐近方差 $\sigma^2_{\beta_r}$ 的相合估计, $z_{1-\alpha/2}$ 是标准正态分布的 $1-\alpha/2$ 上侧分位数, 即 $P(Z\geqslant z_{1-\alpha/2})=\alpha/2$, 其中 Z 表示标准正态随机变量.

1. binning 估计方法的置信区间

由 2.5.1 节给出的参数估计的渐近正态性结果, 见式 (2.5.5), 利用式 (2.6.1) 可构造回归参数 $\beta_r, r=0,1,\cdots,p$ 的置信水平为 $1-\alpha$ 的渐近置信区间为

$$\left[\hat{\beta}_r-\frac{\hat{\sigma}_r}{\sqrt{n}}z_{1-\alpha/2},\hat{\beta}_r+\frac{\hat{\sigma}_r}{\sqrt{n}}z_{1-\alpha/2}\right],\quad r=0,1,\cdots,p.$$

$\hat{\sigma}^2_r$ 是 binning 估计方法得到的渐近方差估计, Sentürk 和 Müller (2006) 证明它们是相合的, 定义如下:

$$\begin{aligned}\hat{\sigma}^2_0=&\left(\sum_{j=1}^m\frac{L_j}{n}\hat{\alpha}^2_{j0}-\hat{\beta}^2_0\right)+\left\{\frac{1}{n}\sum_{j=1}^m\sum_{k=1}^{L_j}\left(\widetilde{Y}^*_{jk}-\hat{\alpha}_{j0}-\sum_{r=1}^p\widetilde{X}^*_{jkr}\hat{\alpha}_{jr}\right)^2\right\}\\&\times\left\{\sum_{j=1}^m\frac{L_j}{n}\left(L_j^{-1}\widetilde{X}^{*\mathrm{T}}_j\widetilde{X}^*_j\right)^{-1}_{(1,1)}\right\},\end{aligned}$$

$$\begin{aligned}\hat{\sigma}^2_r=&\frac{1}{\bar{\widetilde{X}}^2_r}\left[\frac{1}{n}\sum_{j=1}^m\hat{\alpha}^2_{jr}\sum_{k=1}^{L_j}\widetilde{X}^{*2}_{jkr}+\hat{\beta}^2_r\bar{\widetilde{X}}^2_r-2\frac{\hat{\beta}_r}{n}\sum_{j=1}^m\hat{\alpha}_{jr}\sum_{k=1}^{L_j}\widetilde{X}^{*2}_{jkr}+\hat{\beta}^2_rS^2_{\widetilde{X}_r}\right.\\&\left.+\left\{\frac{1}{n}\sum_{j=1}^m\sum_{k=1}^{L_j}\left(\widetilde{Y}^*_{jk}-\hat{\alpha}_{j0}-\sum_{r=1}^p\widetilde{X}^*_{jkr}\hat{\alpha}_{jr}\right)^2\right\}\left\{\sum_{j=1}^m\frac{L_j}{n}\bar{\widetilde{X}}^{*2}_{jr}\left(L_j^{-1}\widetilde{X}^{*\mathrm{T}}_j\widetilde{X}^*_j\right)^{-1}_{(r,r)}\right\}\right],\end{aligned}$$

其中 $S^2_{\widetilde{X}_r}=\dfrac{1}{n-1}\displaystyle\sum_{i=1}^n(\widetilde{X}_{ir}-\bar{\widetilde{X}}_r)^2, r=1,\cdots,p.$

2. 直接估计方法的置信区间

由 2.5.4 节给出的参数估计的渐近正态性结果, 见式 (2.5.5), 利用式 (2.5.14) 可构造回归参数 $\beta_r, r=0,1,\cdots,p$, 的置信水平为 $1-\alpha$ 的渐近置信区间为

$$\left[\hat{\beta}_r-\frac{\hat{\sigma}^*_r}{\sqrt{n}}z_{1-\alpha/2},\hat{\beta}_r+\frac{\hat{\sigma}^*_r}{\sqrt{n}}z_{1-\alpha/2}\right],\quad r=0,1,\cdots,p,$$

其中 $\hat{\sigma}_r^{*2} = e_{r+1,p+1}^{\mathrm{T}}\hat{\Sigma}e_{r+1,p+1}$, 且 Σ 的渐近估计为 $\hat{\Sigma} = \hat{\sigma}^2\hat{\Lambda}^{-1} + \hat{\Lambda}^{-1}\hat{\Xi}\hat{\Lambda}^{-1}$, 这里

$$\hat{\sigma}^2 = \frac{1}{n}\sum_{i=1}^{n}\left(\hat{Y}_i - \hat{\beta}_0 - \sum_{r=1}^{p}\hat{\beta}_r\hat{X}_{ir}\right)^2, \quad \hat{\Lambda} = \frac{1}{n}\sum_{i=1}^{n}\begin{pmatrix}1\\ \hat{X}_i\end{pmatrix}^{\otimes 2} = \frac{1}{n}\hat{\mathcal{X}}^{\mathrm{T}}\hat{\mathcal{X}},$$

$$\begin{aligned}\hat{\Xi} = &\sum_{l=1}^{p}\sum_{s=1}^{p}\frac{\widehat{E}(X_lX_s)}{\bar{\tilde{X}}_l\bar{\tilde{X}}_s}\hat{\pi}_{X_l}\hat{\pi}_{X_s}^{\mathrm{T}}\widehat{\mathrm{Cov}}(\psi_l(U),\psi_s(U))\hat{\beta}_l\hat{\beta}_s\\ &-\sum_{l=1}^{p}\frac{\widehat{E}(X_lY)}{\bar{\tilde{X}}_l\bar{\tilde{Y}}}\left[\hat{\pi}_{X_l}\hat{\pi}_Y^{\mathrm{T}} + \hat{\pi}_Y\hat{\pi}_{X_l}^{\mathrm{T}}\right]\widehat{\mathrm{Cov}}(\psi_l(U),\phi(U))\hat{\beta}_l\\ &+\frac{\frac{1}{n}\sum_{i=1}^{n}\hat{Y}_i^2}{\bar{\tilde{Y}}^2}\hat{\pi}_Y^{\otimes 2}\widehat{\mathrm{Var}}(\phi(U)).\end{aligned}$$

进一步, $\hat{\pi}_{X_l}, \widehat{E}(X_lX_s), \widehat{\mathrm{Cov}}(\psi_l(U),\psi_s(U))$ 定义为

$$\hat{\pi}_{X_l} = \frac{1}{n}\sum_{i=1}^{n}\left(1,\hat{X}_i^{\mathrm{T}}\right)^{\mathrm{T}}\hat{X}_{il}, \quad \hat{\pi}_Y = \frac{1}{n}\sum_{i=1}^{n}\left(1,\hat{X}_i^{\mathrm{T}}\right)^{\mathrm{T}}\hat{Y}_i,$$

$$\widehat{E}(X_lX_s) = \frac{1}{n}\sum_{i=1}^{n}\hat{X}_{il}\hat{X}_{is}, \quad \widehat{E}(X_lY) = \frac{1}{n}\sum_{i=1}^{n}\hat{X}_{il}\hat{Y}_i,$$

$$\widehat{\mathrm{Cov}}(\psi_l(U),\psi_s(U)) = \frac{1}{n}\sum_{i=1}^{n}\hat{\psi}_l(U_i)\hat{\psi}_s(U_i) - 1,$$

$$\widehat{\mathrm{Cov}}(\psi_l(U),\phi(U)) = \frac{1}{n}\sum_{i=1}^{n}\hat{\psi}_l(U_i)\hat{\phi}(U_i) - 1.$$

2.6.2　Bootstrap 重抽样近似方法

由 2.6.1 节可以看出, 基于渐近正态近似方法构造回归参数 $\beta_0,\beta_1,\cdots,\beta_p$ 的置信区间时, 需要估计渐近方差, 而渐近方差估计的结构也非常复杂, 其相合性的大样本性质也很难证明. 为了解决这个问题, Zhang 等 (2014) 在直接估计方法的基础上给出了 wild-bootstrap 置信区间的构造方法, 具体步骤如下.

步骤 1　由直接估计方法, 计算残差 $\hat{\varepsilon}_i = \hat{Y}_i - \hat{\beta}_0 - \sum_{r=1}^{p}\hat{X}_{ir}\hat{\beta}_r$, 其中 $\hat{\beta}_0,\hat{\beta}_1,\cdots,\hat{\beta}_p$ 是由式 (2.5.13) 所定义的最小二乘估计.

步骤 2　对每一个$i = 1,\cdots,n$, 产生 bootstrap 误差 $\varepsilon_i^* = \hat{\varepsilon}_i v_i$, 其中 v_i 是独立的来自于 Bernoulli 分布, 满足 $P\left(v_i = \frac{1-\sqrt{5}}{2}\right) = \frac{5+\sqrt{5}}{10}$, $P\left(v_i = \frac{1+\sqrt{5}}{2}\right) =$

$\frac{5-\sqrt{5}}{10}$. 产生新的响应变量 $\hat{Y}_i^*$, 即

$$\hat{Y}_i^* = \hat{\beta}_0 + \sum_{r=1}^{p} \hat{X}_{ir}\hat{\beta}_r + \varepsilon_i^*.$$

步骤 3 基于 bootstrap 样本 $\{(\hat{Y}_i^*, \hat{X}_i), i = 1, \cdots, n\}$, 由式 (2.5.13) 可获得回归参数的最小二乘估计如下:

$$\begin{pmatrix} \hat{\beta}_0^* \\ \hat{\beta}_1^* \\ \vdots \\ \hat{\beta}_p^* \end{pmatrix} = \left[\frac{1}{n}\sum_{i=1}^{n}\begin{pmatrix} 1 \\ \hat{X}_i \end{pmatrix}^{\otimes 2}\right]^{-1}\left[\frac{1}{n}\sum_{i=1}^{n}\begin{pmatrix} 1 \\ \hat{X}_i \end{pmatrix}\hat{Y}_i^*\right]. \tag{2.6.2}$$

步骤 4 重复 N 次步骤 2 和步骤 3, 可得 β_r 的 N 个估计样本: $\hat{\beta}_{r,k}^*, r = 0, 1, \cdots, p, k = 1, \cdots, N$.

步骤 5 对于 $r = 0, 1, \cdots, p$ 和给定的显著性水平 α, 把步骤 4 中得到的 N 个估计 $\hat{\beta}_{r,1}, \cdots, \hat{\beta}_{r,N}$ 从小到大排序, 然后计算 $\alpha/2$ 样本分位数 $\hat{\beta}_{r,k}^{(\alpha/2)}$ 与 $1-\alpha/2$ 样本分位数 $\hat{\beta}_{r,k}^{(1-\alpha/2)}$. 然后, 可将 $\left[\hat{\beta}_{r,k}^{(\alpha/2)}, \hat{\beta}_{r,k}^{(1-\alpha/2)}\right]$ 来作为 β_r 置信水平为 $1-\alpha$ 的 wild-bootstrap 置信区间.

这里值得注意的是, wild-bootstrap 抽样随机数 v_i 需满足 $E^*(v_i) = 0$, $\text{Var}^*(v_i) = 1$, 以及存在正常数 c, 对所有的 $1 \leqslant i \leqslant n$, 使得 $|v_i| \leqslant c < \infty$. 进一步需要满足, $E^*(\varepsilon_i^*) = 0$, $\text{Var}^*(\varepsilon_i^*) = \hat{\varepsilon}_i^2$, 这里的 $E^*(\cdot)$ 和 $\text{Var}^*(\cdot)$ 的计算是针对随机数本身的分布下计算的 (即计算数学期望或者方差的时候, $\hat{\varepsilon}$ 可看成是常数). Mammen (1993) 建议随机数 v_i 服从上述算法中使用的两点分布, 并且该分布满足三阶矩条件 $E^*(\xi_{ib}^3) = 1$. 该三阶矩条件通常情况下可以数值上改进 wild-bootstrap 的检验效果. 关于 wild-bootstrap 在参数模型下的检验效果, 读者可参考 Stute 等 (1998a).

2.6.3 经验似然方法

本节提供一种构造回归参数置信区间或置信域的备选方法: 经验似然方法. 经验似然方法是由 Owen (1988) 提出的, 它有类似于 bootstrap 重抽样的统计特性. 由于经验似然方法在构造置信域方面有以下几个方面突出的优点, 已经被许多统计学者应用到了各种统计模型的研究中. 例如: ① 经验似然方法不需要预先假定数据所属的分布族, 构造的置信域的形状和大小完全由数据决定, 而且还具有域保持性和变换不变性; ② 经验似然方法具有 Bartlett 纠偏性以及无须构造枢轴统计量; ③ 经验似然方法可避免估计渐近方差或协方差阵, 对于一些半参数模型或数据类型, 渐近方差或协方差阵的结构非常复杂, 很难得到它们的相合估计. 如果利用渐

近正态近似方法构造置信域, 就会严重影响置信域的精度和可靠度, 而经验似然方法正好避免了渐近方差或协方差阵的估计.

为了定义经验对数似然比, 首先引入一些记号. 令 $\beta^* = (\beta_0, \beta_1, \cdots, \beta_p)^{\mathrm{T}}$, $\hat{\mathcal{X}}_i = (1, \hat{X}_i^{\mathrm{T}})^{\mathrm{T}}$, 且令 $\{(\hat{Y}_i, \hat{X}_i = (\hat{X}_{i1}, \cdots, \hat{X}_{ip})^{\mathrm{T}}), i = 1, \cdots, n\}$ 是 2.5.4 节给出的恢复数据集. 定义辅助随机向量如下:

$$\hat{\eta}_i(\beta^*) = \hat{\mathcal{X}}_i\big(\hat{Y}_i - \hat{\mathcal{X}}_i^{\mathrm{T}}\beta^*\big).$$

如果 β^* 是真参数向量, 由 Cui (2008) 或 2.5.4 节的内容, 可以证明 $E[\hat{\eta}_i(\beta^*)] = o(1)$. 因此, 可以定义下面的经验对数似然比函数

$$l_n(\beta^*) = -2\max\left\{\sum_{i=1}^n \log(np_i) \;\middle|\; p_i \geqslant 0, \sum_{i=1}^n p_i = 1, \sum_{i=1}^n p_i\hat{\eta}_i(\beta^*) = 0\right\}. \tag{2.6.3}$$

由 Lagrange 乘子法, 有

$$p_i = \frac{1}{n}\frac{1}{1 + \lambda^{\mathrm{T}}\hat{\eta}_i(\beta^*)}, \quad i = 1, \cdots, n.$$

由上式, 则由式 (2.6.3) 定义的经验对数似然比函数 $l_n(\beta^*)$ 能够重新表示为

$$l_n(\beta^*) = 2\sum_{i=1}^n \log\{1 + \lambda^{\mathrm{T}}\hat{\eta}_i(\beta^*)\},$$

其中 λ 为 Lagrange 乘子, 并满足

$$\frac{1}{n}\sum_{i=1}^n \frac{\hat{\eta}_i(\beta^*)}{1 + \lambda^{\mathrm{T}}\hat{\eta}_i(\beta^*)} = 0.$$

当 $n \to \infty$ 时, 并在一定正则条件下, 容易证明

$$l_n(\beta^*) \xrightarrow{L} \chi^2_{p+1}.$$

由上面的结果, 可以构造参数向量 β^* 的置信域, 即对于任给的 $0 < \alpha < 1$, 存在 c_α 使得 $P(\chi^2_{p+1} > c_\alpha) = \alpha$, 则

$$I_\alpha(\breve{\beta}^*) = \left\{\breve{\beta}^* \in \mathbb{R}^{p+1} \;\middle|\; l_n(\breve{\beta}^*) \leqslant c_\alpha\right\}$$

为参数向量 β^* 的具有渐近置信水平 $1 - \alpha$ 的置信域.

2.7 协变量调整回归模型的变量选择

本节考虑如下的模型:

$$\begin{cases} Y_i = \sum_{r=1}^{p_n} X_{ir}\beta_r + \varepsilon_i, \\ \widetilde{Y}_i = \phi(U_i)Y_i, \\ \widetilde{X}_{ir} = \psi_r(U_i)X_{ir}, \quad i=1,\cdots,n,\ r=1,\cdots,p_n, \end{cases} \tag{2.7.1}$$

其中 $\beta=(\beta_1,\cdots,\beta_{p_n})^{\mathrm{T}}$. 区别于模型 (2.1.7), 本节所考虑模型 (2.7.1) 没有考虑截距项, 最主要的区别是本节考虑的协变量或参数向量的维数随着样本量 n 的增大而发散, 即当 $n\to\infty$ 时, $p_n\to\infty$. 因此, p_n 的下标主要是用于表示协变量或参数向量的维数 p 随着样本量 n 的增大而发散.

在 2.5 节和 2.6 节介绍了协变量调整线性回归模型的估计方法和置信域的构造问题. 本节主要介绍具有发散参数的协变量调整线性回归模型的变量选择问题. 最近, 各种惩罚变量选择方法被发展并成为研究线性回归模型和广义线性回归模型的重要工具. 例如, 桥回归 (bridge regression, Frank and Friedman, 1993), Lasso (Tibshirani, 1996), Elastic-Net (Zou and Hastie, 2005), Adaptive Lasso (Zou, 2006), SCAD (Fan and Li, 2001; Fan and Peng, 2004), 非凹惩罚 M 估计方法 (Li et al., 2011b), 确定独立变量筛选方法 (sure independence screening, Fan and Lv, 2008; Li et al., 2012b).

本节提出 SCAD 惩罚最小二乘方法研究高维协变量调整回归模型, 该方法能够同时进行变量选择和参数估计. 在一定的正则条件下, 所得到的 SCAD 最小二乘估计具有 oracle 性质. 此外, 建议 BIC 准则选择截断参数, 并证明所提出的 BIC 准则能够相合地识别正确模型. 模拟研究和实际数据分析说明了方法的有效性.

2.7.1 估计方法

在稀疏模型中, p_n 维的协变量能够分成两类: 对响应变量具有显著影响的协变量, 对应的系数非零; 对响应变量不显著的协变量, 对应的系数为零. 本节令真实的参数向量表示为 β_0, 对应的 β_0 能够分解成 $\beta_0=(\beta_{I0}^{\mathrm{T}},\beta_{II0}^{\mathrm{T}})^{\mathrm{T}}$, 其中 β_{I0} 是 k_n 维的非零参数向量, β_{II0} 是 p_n-k_n 维的零向量, 而哪些系数非零和哪些系数为零, 提前是不知道的. 如此分解只是为了表达方便. 令 $Y=(Y_1,\cdots,Y_n)^{\mathrm{T}}$, 且 $\mathbb{X}$ 是 $n\times p_n$ 的矩阵. 对应于 β_0 的分解, 记 $\mathbb{X}=(\mathbb{X}_I,\mathbb{X}_{II})$, 其中 $\mathbb{X}_I$ 和 $\mathbb{X}_{II}$ 分别是 $n\times k_n$ 和 $n\times(p_n-k_n)$ 的矩阵. 令 $\mathbb{S}_n=\mathbb{X}^{\mathrm{T}}\mathbb{X}$ 和 $\mathbb{S}_I=\mathbb{X}_I^{\mathrm{T}}\mathbb{X}_I$.

假设可观测数据 $\{(\widetilde{Y}_i,\widetilde{X}_i=(\widetilde{X}_{i1},\cdots,\widetilde{X}_{ip_n})^{\mathrm{T}},U_i),1\leqslant i\leqslant n\}$ 是 i.i.d., 并且来自于模型 (2.7.1). 基于观测数据 $\{(\widetilde{Y}_i,\widetilde{X}_i=(\widetilde{X}_{i1},\cdots,\widetilde{X}_{ip_n})^{\mathrm{T}},U_i),1\leqslant i\leqslant n\}$, 失

真函数的约束条件 (2.1.8), 以及假设 $\{X_r, U, \varepsilon\}$ 相互独立和 $\{Y, U\}$ 相互独立, 利用 2.5.4 节提出的直接估计方法, 可以得到恢复数据集 $\{(\hat{Y}_i, \hat{X}_i), 1 \leqslant i \leqslant n\}$.

基于恢复数据集 $\{(\hat{Y}_i, \hat{X}_i), 1 \leqslant i \leqslant n\}$, 极小化下面的惩罚最小二乘目标函数可对未知参数向量 β 进行估计

$$Q_n(\beta; \lambda, a) = \frac{1}{2}\sum_{i=1}^{n}(\hat{Y}_i - \hat{X}_i^{\mathrm{T}}\beta)^2 + n\sum_{j=1}^{p_n} p_\lambda(|\beta_j|), \tag{2.7.2}$$

其中 $p_\lambda(\cdot)$ 是一个具有截断参数 λ 的预先指定的惩罚函数, 而 λ 能够使用一些数据驱动的方法进行选取, 如交错核实方法 (cross-validation, Craven and Wahba, 1979; Tibshirani, 1996), BIC (Wang et al., 2007; Li et al., 2011b). 对于一个适当的截断参数 λ, 极小化目标函数 (2.7.2), 能够得到 β 的一个稀疏估计, 并完成变量选择.

式 (2.7.2) 中的惩罚函数有许多选择, 本节采用 Fan 和 Li (2001) 提出的 SCAD 惩罚函数, 因为它满足无偏性、连续性和稀疏性, 并且有 oracle 性质. SCAD 惩罚函数有如下的形式

$$p'_\lambda(|\beta|) = \lambda\left\{I(|\beta| \leqslant \lambda) + \frac{(a\lambda - |\beta|)_+}{(a-1)\lambda} I(|\beta| > \lambda)\right\}, \quad a > 2, \tag{2.7.3}$$

其中 a 是另一个截断参数, Fan 和 Li (2001) 从贝叶斯的角度认为 $a = 3.7$ 通常是合适的. β 的 SCAD 惩罚最小二乘估计为

$$\hat{\beta} =: \hat{\beta}(\lambda) = \arg\min_{\beta}\ Q_n(\beta; \lambda). \tag{2.7.4}$$

2.7.2 渐近性质

为了研究在参数维数随着样本量大小发散情况下参数估计的渐近性质, 需要下列的正则条件. 定理的详细证明见 Li 等 (2014b).

(C2.7.1) 存在常数 C_1 和 C_2 使得

$$0 < nC_1 < \lambda_{\min}(\mathbb{S}_n) < \lambda_{\max}(\mathbb{S}_n) < nC_2.$$

(C2.7.2) 所有函数 $g_r(u) = \psi_r(u)f(u)$, $1 \leqslant r \leqslant p_n$, $g_Y(u) = \phi(u)f(u)$, $\psi_r(u)$, $\phi(u)$ 和 $f(u)$ 都是大于一个正常数的可微函数, 并且它们的导数满足: 存在原点的一个邻域 Δ 和常数 $c > 0$, 使得对于任意的 $\delta \in \Delta$, 有

$$\begin{aligned} &|g_r^{(3)}(u+\delta) - g_r^{(3)}(u)| \leqslant c|\delta|, \quad 1 \leqslant r \leqslant p_n, \\ &|g_Y^{(3)}(u+\delta) - g_Y^{(3)}(u)| \leqslant c|\delta|, \\ &|f^{(3)}(u+\delta) - f^{(3)}(u)| \leqslant c|\delta|. \end{aligned}$$

(C2.7.3) 核函数 $K(\cdot)$ 是关于 0 对称的, 并在其支撑集 $[-1, 1]$ 上是连续可微的.

进一步, $\int_{-1}^{1} K(u)\mathrm{d}u = 1$, $\int_{-1}^{1} u^i K(u)\mathrm{d}u = 0, i = 1, 2, 3$.

(C2.7.4) 当 $n \to \infty$ 时, 窗宽 h 在 $O(n^{-\frac{1}{4}}\log n)$ 和 $O(n^{-\frac{1}{8}})$ 的范围内取值.

(C2.7.5) EY 和 EX_r 是有界远离 0 的, 其中 $1 \leqslant r \leqslant p_n$.

(C2.7.6) 对于所有的 p_n,

$$0 < C_1 < \lambda_{\min}\{E(XX^{\mathrm{T}})\} < \lambda_{\max}\{E(XX^{\mathrm{T}})\} < C_2 < \infty.$$

这些通常的条件是很容易被满足的, 并在一些文献中已经被使用. 条件 (C2.7.1) 是研究线性回归模型一个经典的条件. 条件 (C2.7.2) 关于函数 $g_r(\cdot)$ 和 $g_Y(\cdot)$, 以及 U 的密度函数 $f(\cdot)$ 的光滑条件. 条件 (C2.7.3) 用于证明基于核光滑估计的 $\sqrt{n}$ 相合性, 见 Zhu 和 Fang (1996), Härdle 和 Stoker (1989). 条件 (C2.7.4) 给了窗宽选择一个很大的范围, 其中包含了估计未知失真函数的最优窗宽, 同时也表示估计对窗宽不是很敏感的. 条件 (C2.7.5) 是一个特殊条件, 主要是避免分母的估计为 0 或太小而使得估计没有意义, 类似的条件假设见 Sentürk 和 Müller (2005a). 条件 (C2.7.6) 是一个技术条件, 方便定理的证明, 见 Zhu 和 Zhu (2009b).

令 $a_n = \max\{|p'_\lambda(|\beta_{0j}|)| : \beta_{0j} \neq 0\}$, 和 $b_n = \max\{|p''_\lambda(|\beta_{0j}|)| : \beta_{0j} \neq 0\}$, 其中 β_{0j} 是 β_0 的第 j 个分量. 此外, 需要下面关于惩罚函数的条件.

(A2.7.1) $\liminf\limits_{n\to\infty} \liminf\limits_{\theta\to 0+} p'_\lambda(\theta)/\lambda > 0$.

(A2.7.2) $a_n = O(n^{-1/2})$.

(A2.7.3) 当 $n \to \infty$ 时, $b_n \to 0$.

(A2.7.4) 存在常数 C 和 D, 使得当 $\theta_1, \theta_2 > C\lambda$ 时, 满足 $|p''_\lambda(\theta_1) - p''_\lambda(\theta_2)| \leqslant D|\theta_1 - \theta_2|$.

定理 2.7.1 ($\hat\beta$ 的相合性) 假设正则条件(C2.7.1)~ 条件(C2.7.6)成立, 惩罚函数 $p_\lambda(\cdot)$ 满足条件(A2.7.1)~ 条件(A2.7.4). 如果 $n \to \infty$ 时, $p_n^2/n \to 0$, 则有

$$||\hat\beta - \beta_0|| = O_P\{\sqrt{p_n}(n^{-1/2} + a_n)\}.$$

由定理 2.7.1 容易看到, 如果 a_n 满足条件 (A2.7.2), 即 $a_n = O(n^{-1/2})$, 对于适当的截断参数 λ, 则 β 存在一个 $\sqrt{n/p_n}$ 相合估计. 这个收敛速度一致于 Huber (1973) 与 Li 等 (2011b) 研究具有发散参数线性回归模型的 M 估计. 下面的结果显示所得到的 SCAD 惩罚最小二乘估计具有稀疏性 $\hat\beta_{II} = 0$.

为了研究 $\hat\beta$ 的渐近性质, 首先介绍下面一些概念. 令

$$\begin{aligned}\Sigma_\lambda &= \mathrm{diag}\{p''_\lambda(\beta_{01}), \cdots, p''_\lambda(\beta_{0k_n})\},\\ b_n &= \{p'_\lambda(|\beta_{01}|)\mathrm{sgn}(\beta_{01}), \cdots, p'_\lambda(|\beta_{0k_n}|)\mathrm{sgn}(\beta_{0k_n})\}^{\mathrm{T}},\end{aligned}$$

$$\Lambda = \begin{pmatrix} EX_1^2 & EX_1X_2 & \cdots & EX_1X_{p_n} \\ EX_2X_1 & EX_2^2 & \cdots & EX_2X_{p_n} \\ \vdots & \vdots & & \vdots \\ EX_{p_n}X_1 & EX_{p_n}X_2 & \cdots & EX_{p_n}^2 \end{pmatrix}$$

和

$$\Lambda_I = \begin{pmatrix} EX_1^2 & EX_1X_2 & \cdots & EX_1X_{k_n} \\ EX_2X_1 & EX_2^2 & \cdots & EX_2X_{k_n} \\ \vdots & \vdots & & \vdots \\ EX_{k_n}X_1 & EX_{k_n}X_2 & \cdots & EX_{k_n}^2 \end{pmatrix}.$$

此外, 令 $\mathbb{M}$ 是一个 $k_n \times k_n$ 的矩阵, 其中 $\mathbb{M}$ 的第 (r,k) 个元素定义如下, $1 \leqslant r,k \leqslant k_n$

$$\mathbb{M}(r,k) = \beta_r\beta_k E\left\{\frac{EY^2}{(EY)^2}\mathrm{Var}(\phi(U)) + \frac{E(X_rX_k)}{EX_rEX_k}\mathrm{Cov}(\psi_k(U),\psi_r(U)) \right.$$
$$\left. -\frac{EYX_k}{EYEX_k}\mathrm{Cov}(\psi_k(U),\phi(U)) - \frac{EYX_r}{EYEX_r}\mathrm{Cov}(\psi_r(U),\phi(U))\right\}.$$

定理 2.7.2 (Oracle 性质)　假设正则条件(C2.7.1)~条件(C2.7.6)和条件(A2.7.1)~条件(A2.7.4)成立. 如果 $n\to\infty$ 时, $\lambda\to 0$, $\sqrt{n/p_n}\lambda\to\infty$ 和 $p_n^3/n\to 0$, 则以概率趋于 1, 定义于式 (2.7.4) 中 $\sqrt{n/p_n}$ 相合的 SCAD 惩罚最小二乘估计 $\hat\beta = (\hat\beta_I^{\mathrm{T}}, \hat\beta_{II}^{\mathrm{T}})^{\mathrm{T}}$ 满足:

(1) (稀疏性) $\hat\beta_{II} = 0$.

(2) (渐近正态性)

$$\sqrt{n}A_n\Lambda_I^{-1}\big(\Lambda_I+\Sigma_\lambda(\beta_{I0})\big)\left[(\hat\beta_I-\beta_{I0})+\big(\Lambda_I+\Sigma_\lambda(\beta_{I0})\big)^{-1}b_n\right]\xrightarrow{L} N(0,\mathbb{G}), \quad (2.7.5)$$

其中 A_n 是一个 $q\times k_n$ 矩阵使得 $\lambda_{\max}(A_nA_n^{\mathrm{T}})<\infty$ 和 $A_n\Phi A_n^{\mathrm{T}}\to\mathbb{G}$, 且 $\mathbb{G}$ 是一个 $q\times q$ 非负对称矩阵. 协方差阵 Φ 被定义为

$$\Phi = \sigma^2\Lambda_I^{-1} + \mathbb{M}.$$

注 2.7.1　从定理 2.7.2 的渐近正态性可以看出, 非零系数估计的渐近协方差为 $\Phi = \sigma^2\Lambda_I^{-1} + \mathbb{M}$, 其中第一项 $\sigma^2\Lambda_I^{-1}$ 是当数据准确观测, 即 $\phi(\cdot) = 1$ 和 $\psi_r(\cdot) = 1, r = 1,\cdots,p_n$, 惩罚最小二乘估计通常的渐近协方差阵. 第二项 $\mathbb{M}$ 是由于数据失真所引起的损失.

定理 2.7.2 表明即使参数的维数发散时, SCAD 惩罚最小二乘估计的稀疏性、渐近正态性和 oracle 性质仍然是成立的. 当样本量 n 足够大时, $\Sigma_\lambda = 0$ 和 $b_n = 0$, 这

时定理 2.7.2(2) 变为 $\sqrt{n}A_n(\hat{\beta}_I - \beta_{I0}) \to N(0, \mathbb{G})$, 该结论和 β_{I0} 的最小二乘估计有相同的效率, 就好像预先知道 $\beta_{II0} = 0$. 由定理 2.7.1 和定理 2.7.2, 能够发现 SCAD 惩罚最小二乘估计的相合性和渐近正态性分别在条件 $p_n^2/n \to 0$ 和 $p_n^3/n \to 0$ 下成立.

当 $p_n \equiv p$ 固定, 且 $\Sigma_\lambda = 0$ 和 $b_n = 0$ 对足够大的样本量 n 成立, 则结论 (2.7.5) 退化到 Cui (2008) 定理 2.2 的结果: $\sqrt{n}(\hat{\beta}_I - \beta_{I0}) \xrightarrow{L} N(0, \Phi)$.

由定理 2.7.2, 容易发现 $\hat{\beta}_I$ 的渐近协方差阵包含着没有指定的失真函数. 这时估计协方差阵是比较困难的, 进而使得利用定理 2.7.2 对 β 进行统计推断也变得困难. 即使采用 plug-in 方法对协方差阵进行估计, 对于协变量调整回归模型在小样本情况下, 统计推断结果也是很不稳定的 (Cui et al., 2009). 原因是协方差阵的 plug-in 估计卷入了非参数核估计, 具有慢的收敛速度. 为了克服这个问题, 采用类似于 Jin 等 (2003) 提出的再抽样的方法对 $\hat{\beta}_I$ 的渐近分布进行逼近. 特别地, 定义 $\hat{\beta}^*$ 为下面损失函数的极小值:

$$Q_n^*(\beta^*; \lambda, a) = \frac{1}{2}\sum_{i=1}^{n} W_i(\hat{Y}_i - \hat{X}_i^{\mathrm{T}}\beta^*)^2 + n\sum_{j=1}^{p_n} p_\lambda(|\beta_j|), \tag{2.7.6}$$

其中 $W_i (i = 1, \cdots, n)$ 是独立的正的随机变量, 且满足 $E(W_i) = 1$ 和 $\mathrm{Var}(W_i) = 1$. 进一步假定独立于观测数据 $\{\widetilde{Y}_i, \widetilde{X}_i = (\widetilde{X}_{i1}, \cdots, \widetilde{X}_{ip_n})^{\mathrm{T}}, U_i, 1 \leqslant i \leqslant n\}$. 类似于 Jin 等 (2003), 容易证明: 在给定观测数据 $\{\widetilde{Y}_i, \widetilde{X}_i = (\widetilde{X}_{i1}, \cdots, \widetilde{X}_{ip_n})^{\mathrm{T}}, U_i, 1 \leqslant i \leqslant n\}$ 条件下, $\sqrt{n}(\hat{\beta}_I^* - \hat{\beta}_I)$ 的条件分布能够逼近 $\sqrt{n}(\hat{\beta}_I - \beta_{I0})$ 的渐近分布. 注意到在观测数据 $\{\widetilde{Y}_i, \widetilde{X}_i = (\widetilde{X}_{i1}, \cdots, \widetilde{X}_{ip_n})^{\mathrm{T}}, U_i, 1 \leqslant i \leqslant n\}$ 的条件下, $Q_n^*(\beta^*; \lambda, a)$ 中唯一的随机变量是 W_i. 为了逼近 $\hat{\beta}_I$ 的分布, 在固定观测数据 $\{\widetilde{Y}_i, \widetilde{X}_i = (\widetilde{X}_{i1}, \cdots, \widetilde{X}_{ip_n})^{\mathrm{T}}, U_i, 1 \leqslant i \leqslant n\}$ 下, 通过产生随机样本 $(W_1, \cdots, W_n)$, 进而获得 $\hat{\beta}_I^*$ 的集合. 这时 $\hat{\beta}_I$ 的协方差阵能够通过 $\hat{\beta}_I^*$ 的经验样本协方差阵进行逼近, 详细讨论见 Jin 等 (2003) 和 Xu 等 (2010).

2.7.3 算法和 λ 的选取

对于给定的 λ 和 a, 本节把 Fan 和 Li (2001) 提出的局部二次逼近 (local quadratic approximation, LQA) 的算法推广到协变量调整回归模型的 SCAD 惩罚最小二乘估计, 并提出 BIC 准则选择截断参数 λ. 首先介绍 LQA 算法.

假设给定的初始值 β 接近于真实参数 β_0, 如果 β_j 非常接近于 0, 这时设 $\hat{\beta}_j = 0$. 否则惩罚函数被一个二次函数局部逼近为

$$[p_\lambda(|\beta_j|)]' = p_\lambda'(|\beta_j|)\mathrm{sgn}(\beta_j) \approx \{p_\lambda'(|\beta_{0j}|)/|\beta_{0j}|\}\beta_j.$$

换句话说

$$p_\lambda(|\beta_j|) \approx p_\lambda(|\beta_{0j}|) + \frac{1}{2}\{p'_\lambda(|\beta_{0j}|)/|\beta_{0j}|\}(\beta_j^2 - \beta_{0j}^2), \quad \beta_j \approx \beta_{0j}. \tag{2.7.7}$$

这时调整 Newton-Raphson 算法能够找到惩罚最小二乘目标函数 (2.7.2) 的最小值. 特别地, 取初始值为 β^0. 随后, 移去不相关的常量, 惩罚最小二乘目标函数 (2.7.2) 变为

$$Q_n(\beta) = \frac{1}{2}\sum_{i=1}^{n}(\hat{Y}_i - \hat{X}_i^{\mathrm{T}}\beta)^2 + n\sum_{j=1}^{p_n}\frac{p'_\lambda(|\beta_j^0|)}{2|\beta_j^0|}\beta_j^2. \tag{2.7.8}$$

极小化式 (2.7.8) 关于参数向量 β, 有

$$\hat{\beta} = \left(\sum_{i=1}^{n}\hat{X}_i^{\mathrm{T}}\hat{X}_i + nD_\lambda(\beta^0)\right)^{-1}\sum_{i=1}^{n}\hat{X}_i^{\mathrm{T}}\hat{Y}_i, \tag{2.7.9}$$

其中 $D_\lambda(\beta^0)$ 是对角矩阵, 对角元素为 $p'_\lambda(|\beta_j^0|)/|\beta_j^0|, j = 1, \cdots, p_n$. 为了获得 β 的估计, 需要如下的迭代算法:

步骤 1　给定初始值 β^0.

步骤 2　给定 β^k, 通过求解式 (2.7.9) 更新 β 到 β^{k+1}.

步骤 3　迭代步骤 2 直到收敛, 获得 β 的最终估计 $\hat{\beta}$.

在步骤 1 中的初始估计可以通过求解没有惩罚的最小二乘目标函数直接获得. 接下来讨论截断参数 λ 的选取问题, Wang 等 (2007) 提出了一种 BIC 准则选取截断参数, 该准则可以避免广义交错核实方法 (GCV) 所产生的不可忽略的过分拟合问题, 并保证变量选择的相合性. 此处把 BIC 方法应用到协变量调整回归模型, 通过极小化如下的目标函数选取最优的 λ:

$$\mathrm{BIC}(\lambda) = \log(\mathrm{MSE}(\lambda)) + \mathrm{DF}_\lambda(\log(n)/n), \tag{2.7.10}$$

其中$\mathrm{MSE}(\lambda) = \frac{1}{n}\sum_{i=1}^{n}(\hat{Y}_i - \hat{X}_i^{\mathrm{T}}\hat{\beta}_\lambda)^2$, DF_λ 是 β_λ 中非零系数的个数. 令 $\mathcal{M} = \{j_1, \cdots, j_{d_n}\}$ 表示给定候选模型协变量的下标集合, $\mathcal{M}_T$ 表示真实模型, $\mathcal{M}_F$ 表示全模型, 并令 $\mathcal{M}_\lambda$ 表示通过极小化具有截断参数 λ 的 SCAD 惩罚最小二乘目标函数选取的模型. 对于给定的候选模型 $\mathcal{M}$, 首先给定 σ^2 如下的相合估计: $\hat{\sigma}_n^2(\mathcal{M}) = \|\hat{Y} - \hat{\mathbb{X}}\hat{\beta}_{\mathrm{m}}\|^2/n$, 其中 $\hat{\beta}_{\mathrm{m}}$ 是相应的 SCAD 惩罚最小二乘估计. 进一步假设如下条件:

(B1) 对于任给的 $\mathcal{M} \subset \mathcal{M}_F$ 和一些 $\sigma^2(\mathcal{M}) > 0$, 依概率 1 有 $\hat{\sigma}_n^2(\mathcal{M}) \to \sigma^2(\mathcal{M})$;

(B2) 对于任给的 $\mathcal{M} \not\supset \mathcal{M}_T$, 有 $\sigma^2(\mathcal{M}) > \sigma^2(\mathcal{M}_T)$.

上面两个条件是用于研究在模型错误指定情形下, 参数向量的估计问题 (Wang et al., 2007; Liang et al., 2010). 在正则条件和上面两个条件下, 有如下模型选择的相合性.

定理 2.7.3 假设条件(B1)和条件(B2), 以及正则条件(C2.7.1)~ 条件(C2.7.6)成立, 当 $n \to \infty$ 时, 则有

$$P(\mathcal{M}_{\hat{\lambda}} = \mathcal{M}_T) \longrightarrow 1. \tag{2.7.11}$$

2.7.4 模拟研究和实际数据分析

1. 模拟研究

本节通过模拟研究说明 SCAD 惩罚最小二乘估计方法在有限样本下的完成情况, 并和 Lasso 方法进行比较. 在所有模拟和数据分析中核函数取 $K(t) = \dfrac{15}{32}(3 - 7t^2)(1-t^2)I(|t| \leqslant 1)$, 采用去一分量的交错核实方法选取窗宽. 考虑下面的模型:

$$\begin{cases} Y_i = \sum\limits_{r=1}^{p_n} X_{ir}\beta_r + \sigma\varepsilon_i, \\ \widetilde{Y}_i = \phi(U_i)Y_i, \\ \widetilde{X}_{ir} = \psi_r(U_i)X_{ir}, \quad i = 1, \cdots, n, \ r = 1, \cdots, p_n, \end{cases} \tag{2.7.12}$$

其中参数向量取为 $\beta = (3, 1.5, 2, -1, 0.5, 0, \cdots, 0)^{\mathrm{T}}$, 即前面 $k_n = 5$ 个协变量是显著的, 其余的协变量不是显著的, 且参数向量的维数取为 $p_n = \lfloor 5n^{1/2} \rfloor$. 协变量 U 来自于两倍标准偏差的截尾正态分布 $N(6, 1)$. 预测变量和失真函数取为如下形式:

$$\begin{aligned}
&X_1 \sim N(2, 1.144),\ \psi_1(U) = \frac{U+10}{16};\ X_2 \sim N(0.5, 0.25),\ \psi_2(U) = \frac{(U+1)^2}{50}; \\
&X_3 \sim N(1, 1), \qquad \psi_3(U) = \frac{U+3}{9};\ X_4 \sim N(0.8, 0.5), \quad \psi_4(U) = \frac{(U+2)^2}{65}; \\
&X_5 \sim N(1.2, 1.5),\ \psi_5(U) = \frac{U+1}{7}; \\
&X_k \sim N(0, 1), \qquad \psi_k(U) = \frac{U-3}{3}, \quad k = 6, \cdots, p_n; \\
&\varepsilon \sim N(0, 1), \qquad \phi(U) = \frac{(U+3)^2}{82}.
\end{aligned}$$

所有上面的失真函数满足可识别性条件 (2.1.8). 为了比较, 取三种不同的 σ 值, 分别为 $\sigma = 0.1, 0.5$ 和 1, 代表强弱信噪比.

为了避免分母为 0, 引起计算的不稳定, 使用如下的岭估计:

$$\hat{\phi}(u)=\frac{\sum_{i=1}^{n}K\left(\frac{u-U_i}{h}\right)\widetilde{Y}_i}{\sum_{i=1}^{n}K\left(\frac{u-U_i}{h}\right)+n^{-2}}\times\frac{1}{\bar{\widetilde{Y}}+n^{-2}},$$
$$\hat{\psi}_r(u)=\frac{\sum_{i=1}^{n}K\left(\frac{u-U_i}{h}\right)\widetilde{X}_{ir}}{\sum_{i=1}^{n}K\left(\frac{u-U_i}{h}\right)+n^{-2}}\times\frac{1}{\bar{\widetilde{X}}_r+n^{-2}},\quad r=1,\cdots,p_n, \tag{2.7.13}$$

其中 $\bar{\widetilde{Y}}=\frac{1}{n}\sum_{i=1}^{n}\widetilde{Y}_i$, $\bar{\widetilde{X}}_r=\frac{1}{n}\sum_{i=1}^{n}\widetilde{X}_{ir}$, h 是窗宽, $K(\cdot)$ 是核函数. 抽取样本量分别为 $n=100$, 200 和 400 的 1000 个数据集, 即重复 1000 次试验. 对于每种方法, 报告零系数的平均个数, L_2 范数定义的平均估计误差 (EE), 以及模型被欠拟合 (Underfitted)、正确拟合 (Correctly fitted) 和过分拟合 (Overfitted) 的比例. 模拟结果报告在表 2.7.1 中, 其中 "C" 表示零系数被正确估计成零的平均个数, "IC" 表示非零系数被错误估计成零的平均个数. 在过分拟合列中, "1" "2" 和 "$\geqslant 3$" 分别表示选取的模型中包含 1 个、2 个或 2 个以上不重要变量的比列.

表 2.7.1 模型 (2.7.12) 的模拟结果

(n,p_n,σ)	方法	No. of Zeros		EE	Underfitted	Correctly fitted	Overfitted		
		C	IC				1	2	$\geqslant 3$
(100,50,0.1)	oracle	45.000	0.000	0.0376	0.000	1.000	0.000	0.000	0.000
	SCAD	40.495	0.355	0.7653	0.318	0.093	0.033	0.037	0.486
	Lasso	35.446	0.298	0.9342	0.275	0.008	0.013	0.034	0.670
(100,50,0.5)	oracle	45.000	0.000	0.0561	0.000	1.000	0.000	0.000	0.000
	SCAD	40.078	0.445	0.8991	0.398	0.065	0.037	0.036	0.441
	Lasso	35.763	0.362	1.1475	0.324	0.006	0.028	0.035	0.607
(100,50,1)	oracle	45.000	0.000	0.1231	0.000	1.000	0.000	0.000	0.000
	SCAD	39.840	0.632	1.3220	0.533	0.020	0.022	0.031	0.381
	Lasso	36.727	0.553	1.6092	0.471	0.004	0.011	0.023	0.491
(200,70,0.1)	oracle	65.000	0.000	0.0132	0.000	1.000	0.000	0.000	0.000
	SCAD	63.161	0.116	0.2847	0.116	0.488	0.027	0.027	0.332
	Lasso	56.726	0.053	0.3423	0.053	0.036	0.075	0.076	0.760
(200,70,0.5)	oracle	65.000	0.000	0.0258	0.000	1.000	0.000	0.000	0.000
	SCAD	63.040	0.205	0.3958	0.198	0.378	0.027	0.031	0.347
	Lasso	57.952	0.111	0.4865	0.110	0.036	0.069	0.082	0.703

续表

(n,p_n,σ)	方法	No. of Zeros		EE	Underfitted	Correctly fitted	Overfitted		
		C	IC				1	2	$\geqslant 3$
(200,70,1)	oracle	65.000	0.000	0.0528	0.000	1.000	0.000	0.000	0.000
	SCAD	62.919	0.360	0.6780	0.343	0.211	0.060	0.052	0.300
	Lasso	59.324	0.270	0.7826	0.265	0.028	0.073	0.092	0.542
(400,100,0.1)	oracle	95.000	0.000	0.0056	0.000	1.000	0.000	0.000	0.000
	SCAD	94.022	0.024	0.0878	0.024	0.667	0.032	0.032	0.235
	Lasso	88.648	0.000	0.1687	0.000	0.072	0.109	0.134	0.685
(400,100,0.5)	oracle	95.000	0.000	0.0109	0.000	1.000	0.000	0.000	0.000
	SCAD	94.309	0.048	0.1949	0.048	0.754	0.028	0.027	0.134
	Lasso	89.915	0.002	0.2414	0.002	0.077	0.169	0.153	0.599
(400,100,1)	oracle	95.000	0.000	0.0242	0.000	1.000	0.000	0.000	0.000
	SCAD	94.366	0.104	0.2654	0.104	0.620	0.064	0.042	0.145
	Lasso	91.139	0.034	0.4140	0.034	0.107	0.203	0.160	0.496

表 2.7.1 说明了如下的结果.

(1) 随着样本量的增大或者噪声水平的减小, 所提出的 SCAD 惩罚最小二乘估计的估计误差快速逼近 oracle 估计的估计误差. 因此, 在估计误差方面, SCAD 方法完成的比 Lasso 要好.

(2) SCAD 方法正确识别真实模型的比例要比 Lasso 方法要好. 对于过分拟合情况, SCAD 方法包含 1 个或 2 个不重要变量的比例要大, 而 Lasso 方法包含 3 个或 3 个以上不重要的变量的比例要大.

(3) 即使对 $\sigma=0.1$ 和 $n=400$ 的情况, Lasso 方法仍然严重地有过分拟合的情况发生, 而 SCAD 方法却相对要好很多.

一般而言, 根据零系数的平均个数, 平均估计误差, 模型欠拟合、正确拟合和过分拟合的比例, SCAD 方法比 Lasso 方法选取了更小的模型并得到了更加精确的模型. 因此, SCAD 方法完成的要比 Lasso 方法更好.

2. 实际数据分析

Harrison 和 Rubinfeld (1978) 利用各种方法对 Boston 住房数据进行分析, 研究和估计空气质量对住房价格的需求. 他们发布了 1970 年 Boston 标准都市统计区 (SMSA) 住房价格的研究报告. Boston 数据集包含 506 个样本和 14 个非常量的独立变量. 这些变量包括: 平均在 1000 美元的私人住房中间价 (MEDV), 二氧化氮的浓度 (NOX, 单位是 pphm), 平均每人所住房间数 (RM), 1940 年前业主单位所占的比例 (AGE), 城镇黑人比例 (B), 地位较低人士的比例 (LSTAT), 城镇人均犯罪率 (CRIM), 规划住宅用地超过 25000 平方英尺的比例 (ZN), 城镇每英亩非零售业的比例 (INDUS), 所有财产价值的每 10000 美元的税率 (TAX), 城镇师生比

例 (PTRATIO), 查尔斯河 (CHAS, 虚拟变量, 如果是大片土地则为 1, 否则为 0), 到达 Boston 就业中心的加权距离 (DIS), 到达径向公路的系数 (RAD). Harrison 和 Rubinfeld (1978) 的表 IV 中对每个变量的详细定义进行了描述. 该数据研究的主要目的是研究各种变量对 Boston 住房价格的影响. 该数据集已经被用于协变量调整回归模型的研究, 如 Sentürk 和 Müller (2005b), Zhang 等 (2012a, 2012b). Sentürk 和 Müller (2005b) 发现教育水平变量 LSTAT 可能是一个影响其他变量的混淆变量, 并且发现变量 CRIM 对住房价格有重要的影响. Zhang 等 (2012b) 也使用 LSTAT 作为一个混淆变量对该数据集进行了分析.

相应于 Harrison 和 Rubinfeld (1978), 对住房价格 MEDV 作对数变换, 然后作为响应变量, 空气污染和其他 11 个变量作为协变量, 它们可能影响着住房价格. 类似于 Sentürk 和 Müller (2005b), 此处也把 LSTAT 作为混淆变量 U. 除了变量 AGE, CHAS 和 ZN, 其他所有变量都需要使用混淆变量 U 进行修正. 模型如下:

$$\begin{aligned}\log(\text{MEDV}) = &\beta_0 + \beta_1(\text{RM})^2 + \beta_2\text{AGE} + \beta_3\log(\text{DIS}) + \beta_4\log(\text{RAD}) + \beta_5\text{TAX}\\ &+\beta_6\text{PTRATIO} + \beta_7(\text{B} - 0.63)^2 + \beta_8\text{CRIM}\\ &+\beta_9\text{ZN} + \beta_{10}\text{INDUS} + \beta_{11}\text{CHAS} + \beta_{12}\text{NOX}^k + \varepsilon,\end{aligned}$$

其中根据数值最优化可得变量 NOX 的最优的幂为 $k = 2$. 既然在模型中参数的个数是比较多的, 并假定依赖于样本量大小, 所提出的 SCAD 惩罚最小二乘估计可用于 Boston 住房数据, 且 BIC 准则可用于选取截断参数 λ. 采用 2.7.2 节提出的再抽样方法计算非零系数估计的标准差, 即从具有参数为 1 的指数分布中随机抽取 1000 个随机变量 W. 表 2.7.2 中报告了估计的系数和它们相应的标准差. 容易看到 SCAD 和 Lasso 两种方法都选出了最简单的模型, 两种方法都是同时进行变量选择和参数估计比较吸引人的方法.

表 2.7.2　Boston 住房数据的估计和计算结果

变量	LS	LASSO	SCAD
β_0	3.3730(3.4169)	2.9682(3.1491)	2.9804(3.4907)
RM^2	0.0057(0.0427)	0.0071(0.0444)	0.0070(0.0464)
AGE	0.0013(0.0107)	0.0012(0.0104)	0.0012(0.0098)
log(DIS)	−0.0283(0.8309)	0(−)	0(−)
log(RAD)	0.0693(0.3263)	0.0595(0.3229)	0.0557(0.3238)
TAX	−0.0002(0.0022)	−0.0003(0.0022)	−0.0003(0.0023)
PTRATIO	−0.0350(0.0930)	−0.0237(0.0917)	−0.0240(0.1017)
$(\text{B}-0.63)^2$	0.0004(0.0029)	0.0005(0.0030)	0.0005(0.0030)
CRIM	−0.0094(0.0419)	−0.0090(0.0424)	−0.0090(0.0424)
ZN	−0.0005(0.0096)	0(−)	0(−)

续表

变量	LS	LASSO	SCAD
INDUS	0.0014(0.0346)	0(−)	0(−)
CHAS	0.1002(0.8303)	0.0775(0.8477)	0.0583(0.8605)
$(\mathrm{NOX})^2$	−0.3928(3.1686)	0(−)	0(−)

2.8 模型检验

2.8.1 引言

在许多实际问题的研究中, 需要对回归函数进行拟合优度检验, 即检验响应变量 Y 与协变量 X 之间是否可以用事先给定的模型进行数据拟合. 本节研究协变量调整回归模型的线性检验问题, 考虑如下的协变量调整回归模型:

$$\begin{cases} Y = g(X) + \varepsilon, \\ \tilde{Y} = \phi(U)Y, \\ \tilde{X}_r = \psi_r(U)X_r, \quad r = 1, \cdots, p, \end{cases} \tag{2.8.1}$$

其中 Y 是不可观测的响应变量, 而观测到的是被混淆变量 U 的未知失真函数 $\phi(\cdot)$ 污染后的变量 $\widetilde{Y}$. $X = (X_1, \cdots, X_p)^{\mathrm{T}}$ 是不可观测的连续型协变量向量, 观测到的是受到未知失真函数 $\psi_r(\cdot)$ $(r = 1, \cdots, p)$ 污染以后的协变量向量 $\widetilde{X} = (\widetilde{X}_1, \cdots, \widetilde{X}_p)^{\mathrm{T}}$. $g(\cdot)$ 是响应变量 Y 关于协变量向量 X 的未知光滑函数, ε 是独立同分布的随机误差, 且有 $E(\varepsilon) = 0$ 和 $\mathrm{Var}(\varepsilon) = \sigma^2$. 未知的失真函数 $\phi(\cdot)$ 和 $\psi_1(\cdot), \cdots, \psi_p(\cdot)$ 允许假定为变量 U 的一般光滑函数, 并且满足约束条件 (2.1.8).

对于模型 (2.8.1), 一个重要的问题是检验回归函数 $E(Y|X) = g(X)$ 是否可以用事先给定的一个参数模型进行拟合, 记 $\mathcal{M} = \{g(\cdot, \beta) : \beta \in \Theta \subset \mathbb{R}^q\}$ 是一个给定的参数函数族, 其中 $\beta = (\beta_1, \cdots, \beta_q)^{\mathrm{T}}$ 是未知的 q 维真实参数, 则考虑如下的假设检验问题:

$$H_0: \text{存在某个参数向量 } \beta \text{和某个回归函数} g(\cdot,\cdot) \in \mathcal{M}, \text{有}$$
$$g(X) = g(X, \beta);$$
$$H_1: \text{对任意的参数向量 } \beta \text{ 和任意的回归函数 } g(\cdot,\cdot), \text{都有}$$
$$g(X) \neq g(X, \beta).$$

模型的拟合优度检验问题一直是统计学家考虑的热点问题, 当 (Y, X) 可观测时, 一些统计学者对上述假设检验问题进行了研究. 例如, 若 $g(X, \beta)$ 是一个线性回归模型时, Azzalini 和 Bowman (1993) 使用光滑后的残差, 并提出了一个伪似

然比 (pseudo-likelihood ratio) 检验统计量, 对参数模型与非参数模型之间的距离进行度量. Lin 等 (2002) 通过计算一定方向下的累积残差和, 提出了主观信息化 (objective and informative) 的模型检验方法. 对于更一般的非线性参数模型, Härdle 和 Mammen (1993) 与 Stute 等 (1998a) 提出了一种模型检验方法, 该方法主要是基于非参数模型拟合与参数模型拟合之间的平方距离, 以及由残差表示的回归变量的经验过程所定义的检验统计量, 对模型进行检验. Stute 等 (1998b) 进一步提出一种新的检验方法, 该方法很容易计算检验统计量的 p 值. Dette (1999) 提出了一种检验方法, 是通过在原假设和备择假设下分别计算方差, 然后对两个方差做差来构造检验统计量. 假设模型误差服从正态分布时, Fan (1996) 与 Fan 和 Huang (2001) 提出了自适应的 Neyman 阈值检验方法, Li 等 (2014a) 将自适应的 Neyman 阈值检验方法推广到单指标模型的研究, 对联系函数是否是线性结构进行假设检验. Fan 等 (2001) 提出了广义似然比 (generalized likelihood ratio, GLR) 统计量, 并证明了统计量有渐近 Wilks 现象. 各种非参数统计方法在模型检验中的详细讨论及其应用, 请参考 Hart (1997).

2.8.2 检验统计量及其讨论

本节主要检验回归函数是否是线性结构, 即考虑如下的假设检验问题:

$$\begin{aligned} &H_0: \text{存在参数向量 } \beta=(\beta_0,\beta_1,\cdots,\beta_p)^{\mathrm{T}}, \text{使得} \\ &\quad g(X)=g(X,\beta)=\beta_0+\beta_1X_1+\cdots+\beta_pX_p, \\ &H_1: \text{对任意的参数向量 } \beta=(\beta_0,\beta_1,\cdots,\beta_p)^{\mathrm{T}}, \text{有} \\ &\quad g(X)\neq g(X,\beta)\neq \beta_0+\beta_1X_1+\cdots+\beta_pX_p. \end{aligned}$$

假设模型误差为 ε_i, 不可观测的样本为 $\{Y_i, X_i=(X_{i1},\cdots,X_{ip})^{\mathrm{T}}\}, i=1,\cdots,n$. 针对上述假设检验问题, 如果 $\{Y_i, X_i=(X_{i1},\cdots,X_{ip})^{\mathrm{T}}\}$ 可观测, Stute 等 (2008a) 构造了如下中心化的经验过程

$$\mathcal{T}(t)=\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\left[l(X_i)-\bar{l}(X)\}\right]I\{\varepsilon_i\leqslant t\},\quad t\in\mathbb{R}, \tag{2.8.2}$$

其中 $l(\cdot)$ 是一个事先指定的关于协变量 X 的实值权函数, $\bar{l}(X)=\dfrac{1}{n}\sum\limits_{i=1}^{n}l(X_i)$. 式 (2.8.2) 所定义的中心化经验过程的检验方法有一些优良性质: ①该检验方法对于全局的备择假设是相合的; ②当局部备择假设以 $n^{-1/2}$ 的速度逼近于原假设 H_0 时, 该检验方法也能够很好的进行检测; ③给定样本值, $\mathcal{T}(t)$ 在 Skorokhod 空间 $D[-\infty,+\infty]$ 上是右连续有界, 左极限存在的函数. 因此, 当 $l(X)$ 的二阶矩存在时, 可以证明中心化的经验过程 $\mathcal{T}(t)$ 依分布收敛到零均值的布朗运动. 基于这些优点,

Zhang 等 (2015b) 将统计量 $\mathcal{T}(t)$ 用于解决协变量调整回归模型的检验问题. 另外, 在式 (2.8.2) 中, 需要选择权函数 $l(\cdot)$, 而 $l(\cdot)$ 的选择取决于偏离待检验零模型的情况. 在实际问题中, 根据具体问题, 可以选择线性函数、二次函数、正余弦函数和指数函数等.

式 (2.8.2) 所定义的中心化经验过程不能直接用于上述的假设检验问题, 因为在式 (2.8.2) 中, $\{(Y_i, X_i), i = 1, \cdots, n\}$ 是不可观测的, 而观测到的是受到混淆变量的失真函数污染以后的变量 $\{(\widetilde{Y}_i, \widetilde{X}_i), i = 1, \cdots, n\}$. 另外, 模型误差 ε_i 也是未知的. 为了构造检验统计量, 首先需要把这些未知量估计出来. 具体步骤为: ① 利用 2.5.4 节提出的直接估计方法, 得到恢复数据集 $\{(\hat{Y}_i, \hat{X}_i), 1 \leqslant i \leqslant n\}$; ②基于恢复数据集 $\{(\hat{Y}_i, \hat{X}_i), 1 \leqslant i \leqslant n\}$, 利用最小二乘方法, 获得回归参数的最小二乘估计为 $\hat{\beta}_0, \hat{\beta}_1, \cdots, \hat{\beta}_p$; ③计算残差 $\hat{\varepsilon}_i = \hat{Y}_i - \hat{\beta}_0 - \hat{\beta}_1 \hat{X}_{i1} - \cdots - \hat{\beta}_p \hat{X}_{ip}$. 有了这些估计量, 构造如下的经验过程检验统计量

$$\mathcal{T}_n(t) = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \left[l(\hat{X}_i) - \bar{l}(\hat{X}) \right] I\{\hat{\varepsilon}_i \leqslant t\}, \quad t \in \mathbb{R}, \tag{2.8.3}$$

其中 $\bar{l}(\hat{X}) = \dfrac{1}{n} \sum_{i=1}^{n} l(\hat{X}_i)$.

接下来, 讨论 $\mathcal{T}_n(t)$ 的一些大样本渐近性质. 首先介绍一些概念, 令 $d(X, \beta) = \beta_0 + \beta_1 X_1 + \cdots + \beta_p X_p - g(X)$, $F_\varepsilon(\cdot) = P(\varepsilon \leqslant \cdot)$, $F'_\varepsilon(t) = \dfrac{\partial F_\varepsilon(t)}{\partial t}$, 以及 $l'_r(X) = \dfrac{\partial l(X)}{\partial X_r}$, $r = 1, \cdots, p$, 其中 $X = (X_1, \cdots, X_p)^{\mathrm{T}}$. 进一步定义

$$\begin{aligned}
\kappa_r &= E\left[X_r l'_r(X) \{F_\varepsilon(t + d(X, \beta)) - E[F_\varepsilon(t + d(X, \beta))]\}\right], \\
\iota_r &= E\Big[X_r \{l(X) - E[l(X)]\} F'_\varepsilon(t + d(X, \beta)) \beta_r\Big], \\
\mu_Y &= E\left[Y \{l(X) - E[l(X)]\} F'_\varepsilon(t + d(X, \beta))\right], \\
\varsigma &= E\left[\{l(X) - E[l(X)]\} F'_\varepsilon(t + d(X, \beta)) (1, X^{\mathrm{T}})^{\mathrm{T}}\right].
\end{aligned}$$

对于 $r = 1, \cdots, p$, 假设权函数 $l(X)$, 及其偏导数 $l'_r(X) = \dfrac{\partial l(X)}{\partial X_r}$ 连续有界, 在一些条件下, 复杂推导可以得到

$$\begin{aligned}
\mathcal{T}_n(t) = &\frac{1}{\sqrt{n}} \sum_{i=1}^{n} \left[l(X_i) - E\{l(X)\}\right] \left[I\{\varepsilon_i \leqslant t + d(X_i, \beta)\} - E\{F_\varepsilon(t + d(X, \beta))\}\right] \\
&+ \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \left\{ \sum_{r=1}^{p} \left[\widetilde{X}_{ir} - X_{ir}\right] \frac{(\kappa_r + \iota_r)}{E[X_r]} - \left[\widetilde{Y}_i - Y_i\right] \frac{\mu_Y}{E[Y]} \right\} \\
&+ \sqrt{n} \varsigma^{\mathrm{T}} \Lambda^{-1} W_n(\beta) + o_P(1),
\end{aligned}$$

其中 $W_n(\beta) = (W_n(\beta_0), W_n(\beta_1), \cdots, W_n(\beta_p))^{\mathrm{T}}$, 并且

$$\begin{aligned} W_n(\beta_0) =& \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \left[Y_i - (1, X_i^{\mathrm{T}})\beta\right] + \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \left[\widetilde{Y}_i - Y_i\right], \\ W_n(\beta_r) =& \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \left[Y_i - (1, X_i^{\mathrm{T}})\beta\right] X_{ir} + \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \left[\widetilde{Y}_i - Y_i\right] \frac{E[YX_r]}{E[Y]} \\ & - \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \sum_{s=1}^{p} \left[\widetilde{X}_{is} - X_{is}\right] \frac{E[X_s X_r]}{E[X_s]} \beta_r, \quad r = 1, \cdots, p. \end{aligned}$$

下面对所得结论给一些解释. $\mathcal{T}_n(t)$ 的渐近表达式中第一项可分解为下面两项, 即

$$\frac{1}{\sqrt{n}} \sum_{i=1}^{n} \left[I\{\varepsilon_i \leqslant t + d(X_i, \beta)\} - F_\varepsilon(t + d(X_i, \beta))\right] \left[l(X_i) - E\{l(X)\}\right] \tag{2.8.4}$$

和

$$\frac{1}{\sqrt{n}} \sum_{i=1}^{n} \left[F_\varepsilon(t + d(X_i, \beta)) - E\left[F_\varepsilon(t + d(X, \beta))\right]\right] \left[l(X_i) - E\{l(X)\}\right]. \tag{2.8.5}$$

根据模型误差 ε_i 和协变量 X_i 的独立性, 可看出式 (2.8.4) 是一个零均值的紧过程. 那么, 当原假设 H_0 成立时, $d(X, \beta) = 0$ 以及 $F_\varepsilon(t + d(X_i, \beta)) = F_\varepsilon(t)$, 此时式 (2.8.5) 也是一个零均值的紧过程. 因此, 在原假设 H_0 成立时, $\mathcal{T}_n(t)$ 是零均值的紧过程并渐近收敛到一个高斯过程. 进一步, 在备择假设 H_1 下, 如果

$$E\left[\left[l(X) - E\{l(X)\}\right] F_\varepsilon(t + d(X, \beta))\right] \neq 0. \tag{2.8.6}$$

当 $n \to \infty$ 时, $|\mathcal{T}_n(t)|$ 的取值将依概率 1 趋于无穷大, 即该检验过程 $\mathcal{T}_n(t)$ 在备择假设 H_1 下有渐近趋于 1 的功效函数. 因此, 所构造的检验过程是一个相合的检验过程.

下面讨论该检验过程的局部渐近功效问题. 考虑如下的备择假设:

$$\begin{aligned} & H_{1n}': \text{存在参数向量 } \beta = (\beta_0, \beta_1, \cdots, \beta_p)^{\mathrm{T}}, \text{使得} \\ & g(X) = \beta_0 + \beta_1 X_1 + \cdots + \beta_p X_p + n^{-1/2} s(X), \end{aligned}$$

其中 $s(X)$ 是关于 X 指定的函数. 局部备择假设 H_{1n}' 以 $n^{-1/2}$ 的速度收敛到原假设 H_0. 在局部备择假设 H_{1n}' 下, 有

$$F_\varepsilon(t + d(X, \beta)) = F_\varepsilon\left(t - n^{-1/2} s(X)\right) = F_\varepsilon(t) - n^{-1/2} F_\varepsilon'(t) s(X) + O_P(n^{-1}).$$

这时式 (2.8.5) 变为

$$
\begin{aligned}
&-\frac{1}{n}\sum_{i=1}^{n}\left[l(X_i)-E\{l(X)\}\right]\left[s(X_i)-E\{s(X)\}\right]F'_{\varepsilon}(t)+O_P(n^{-1/2})\\
&\xrightarrow{P}-\mathrm{Cov}\left(l(X),s(X)\right)F'_{\varepsilon}(t).
\end{aligned}
$$

容易看出, 当原假设 H_0 成立时, $s(X)\equiv 0$, $\mathrm{Cov}\left(l(X),s(X)\right)\equiv 0$. 反过来, 在局部备择假设 H'_{1n} 下, 检验过程 $\mathcal{T}_n(t)$ 可以进一步表示为

$$
\begin{aligned}
\mathcal{T}_n(t)=&\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\left[l\left(X_i\right)-E\{l(X)\}\right]\left[I\left\{\varepsilon_i\leqslant t+d\left(X_i,\beta\right)\right\}-F_{\varepsilon}\left(t+d\left(X_i,\beta\right)\right)\right]\\
&+\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\left\{\sum_{r=1}^{p}\left[\widetilde{X}_{ir}-X_{ir}\right]\frac{(\kappa_r+\iota_r)}{E[X_r]}-\left[\widetilde{Y}_i-Y_i\right]\frac{\mu_Y}{E[Y]}\right\}\\
&+\sqrt{n}\varsigma^{\mathrm{T}}\Lambda^{-1}W_n\left(\beta\right)-\mathrm{Cov}\left(l(X),s(X)\right)F'_{\varepsilon}(t)+o_P(1).
\end{aligned}
$$

由上面的讨论, 可知在局部备择假设 H'_{1n} 下, 检验过程 $\mathcal{T}_n(t)$ 依概率 1 渐近收敛到均值为 $-\mathrm{Cov}\left(l(X),s(X)\right)F'_{\varepsilon}(t)$ 的高斯过程. 说明当局部备择假设以 $n^{-1/2}$ 的速度逼近于原假设 H_0 时, 该检验方法也能够很好地进行检测.

2.8.3 Bootstrap 方法

2.8.2 节讨论了检验统计的构造, 并从理论上研究了检验统计量的优点. 但是在实际应用时, 检验统计量 $\mathcal{T}_n(t)$ 不能直接用来进行拟合优度检验, 因为 $\mathcal{T}_n(t)$ 的渐近分布很难得到, 如此就很难获得检验统计量的临界值. 为了解决实际应用的问题, 构造如下的检验统计量, 并提出 wild-bootstrap 算法来近似检验统计量 $\mathcal{T}_n$ 在原假设 H_0 下的渐近分布.

$$
\mathcal{T}_n=\frac{1}{n}\sum_{j=1}^{n}\mathcal{T}_n^2\left(\hat{\varepsilon}_j\right)=\frac{1}{n^2}\sum_{j=1}^{n}\left[\sum_{i=1}^{n}\left\{l(\hat{X}_i)-\bar{l}(\hat{X})\right\}I(\hat{\varepsilon}_i\leqslant\hat{\varepsilon}_j)\right]^2. \tag{2.8.7}
$$

下面给出 wild-bootstrap 算法的具体步骤.

步骤 1 由直接估计方法, 计算残差 $\hat{\varepsilon}_i=\hat{Y}_i-\hat{\beta}_0-\sum_{r=1}^{p}\hat{X}_{ir}\hat{\beta}_r$, 其中 $\hat{\beta}_0,\hat{\beta}_1,\cdots,\hat{\beta}_p$ 是假定原假设 H_0 成立, 由式 (2.5.13) 所定义的最小二乘估计. 选定权函数 $l(\cdot)$, 计算由式 (2.8.7) 定义的检验统计量 $\mathcal{T}_n$.

步骤 2 对每一个 $i=1,\cdots,n$, 产生 bootstrap 误差 $\varepsilon_i^*=\hat{\varepsilon}_i v_i$, 其中 v_i 是独立的来自于 Bernoulli 分布, 满足 $P\left(v_i=\frac{1-\sqrt{5}}{2}\right)=\frac{5+\sqrt{5}}{10}$, $P\left(v_i=\frac{1+\sqrt{5}}{2}\right)=$

$\frac{5-\sqrt{5}}{10}$. 产生新的响应变量 $\hat{Y}_i^*$, 即

$$\hat{Y}_i^* = \hat{\beta}_0 + \sum_{r=1}^{p} \hat{X}_{ir}\hat{\beta}_r + \varepsilon_i^*.$$

步骤 3　基于 bootstrap 样本 $\{(\hat{Y}_i^*, \hat{X}_i), i = 1, \cdots, n\}$, 由式 (2.6.2) 可获得回归参数的最小二乘估计 $\hat{\beta}_0^*, \hat{\beta}_1^*, \cdots, \hat{\beta}_p^*$. 基于所得到 bootstrap 最小二乘估计, 计算 bootstrap 残差 $\hat{\varepsilon}_i^{*b}$, 进而计算 bootstrap 检验统计量 $\mathcal{T}_n^{*b}$ 如下

$$\mathcal{T}_n^{*b} = \frac{1}{n^2}\sum_{j=1}^{n}\left[\sum_{i=1}^{n}\left\{l(\hat{X}_i) - \bar{l}(\hat{X})\right\} I(\hat{\varepsilon}_i^{*b} \leqslant \hat{\varepsilon}_j^{*b})\right]^2.$$

步骤 4　重复 N 次步骤 2 和步骤 3, 可得 N 个检验统计量的估计样本: $\mathcal{T}_{n,k}^{*b}$, $k = 1, \cdots, N$.

步骤 5　可用 $\mathcal{T}_{n,k}^{*b}, k = 1, \cdots, N$ 的经验分布函数来逼近检验统计量 $\mathcal{T}_n$ 的渐近分布, 或计算 p 值为

$$\hat{p} = \frac{1}{N}\sum_{k=1}^{N} I\left\{\mathcal{T}_{n,k}^{*b} \geqslant \mathcal{T}_n\right\}.$$

2.8.4　模拟研究

本节报告一个模拟研究说明 wild-bootstrap 逼近方法在有限样本下的表现. 使用 Epanechnikov 核函数 $K(t) = 0.75(1-t^2)I(|t| \leqslant 1)$, 样本容量取 $n = 300, 500$, 重复模拟 500 次. 数据集来自于如下模型:

$$\begin{cases} Y = 2X_1 + 3X_2 + C_0\exp(0.5X_2) + \varepsilon, \\ \widetilde{Y} = \phi(U)Y, \\ \widetilde{X}_r = \psi_r(U)X_r, \quad r = 1, 2, \end{cases} \tag{2.8.8}$$

其中 C_0 选取为 0, 0.20, 0.40, 0.60, 0.80. 不难看出, 当 $C_0 = 0$ 时, 模型 (2.8.8) 变成了原假设 H_0 下的线性模型. 当 $C_0 \neq 0$ 时, 模型 (2.8.8) 就是备择假设 H_1 下的非线性模型. X_1, X_2, ε 相互独立, 且 X_1, X_2 来自于区间 $[1,2]$ 上的均匀分布, 而 ε 来自于正态分布 $N(0, 0.15^2)$. 失真函数取为 $\phi(U) = 1 + 0.2\cos(2\pi U)$, $\psi_1(U) = 1 + 0.3\cos(2\pi U)$, $\psi_2(U) = 1 + 0.2\left(U^2 - \frac{1}{3}\right)$, 协变量 U 独立产生于区间 $[0,1]$ 上均匀分布. 此外, wild-bootstrap 重抽样次数 N 选取为 1000 次. 权函数的选取为 $l(X) = s(X) = \exp(0.5X_2)$. 此时, $\mathrm{Cov}(l(X), s(X)) = \mathrm{Var}(l(X))$ 是一个非零的正值. 利用直接估计方法获取恢复数据时, 窗宽 h 的选取采用 Silverman (1986)

提出的拇指准则, 取为 $\hat{h}=\hat{\sigma}_U n^{-1/3}$, 其中 $\hat{\sigma}_U$ 表示样本 $U_1,\cdots,U_n$ 的样本标准差. 将不同 C_0 取值下的模拟结果报告在图 2.8.1 中, 其中 α 表示显著性水平.

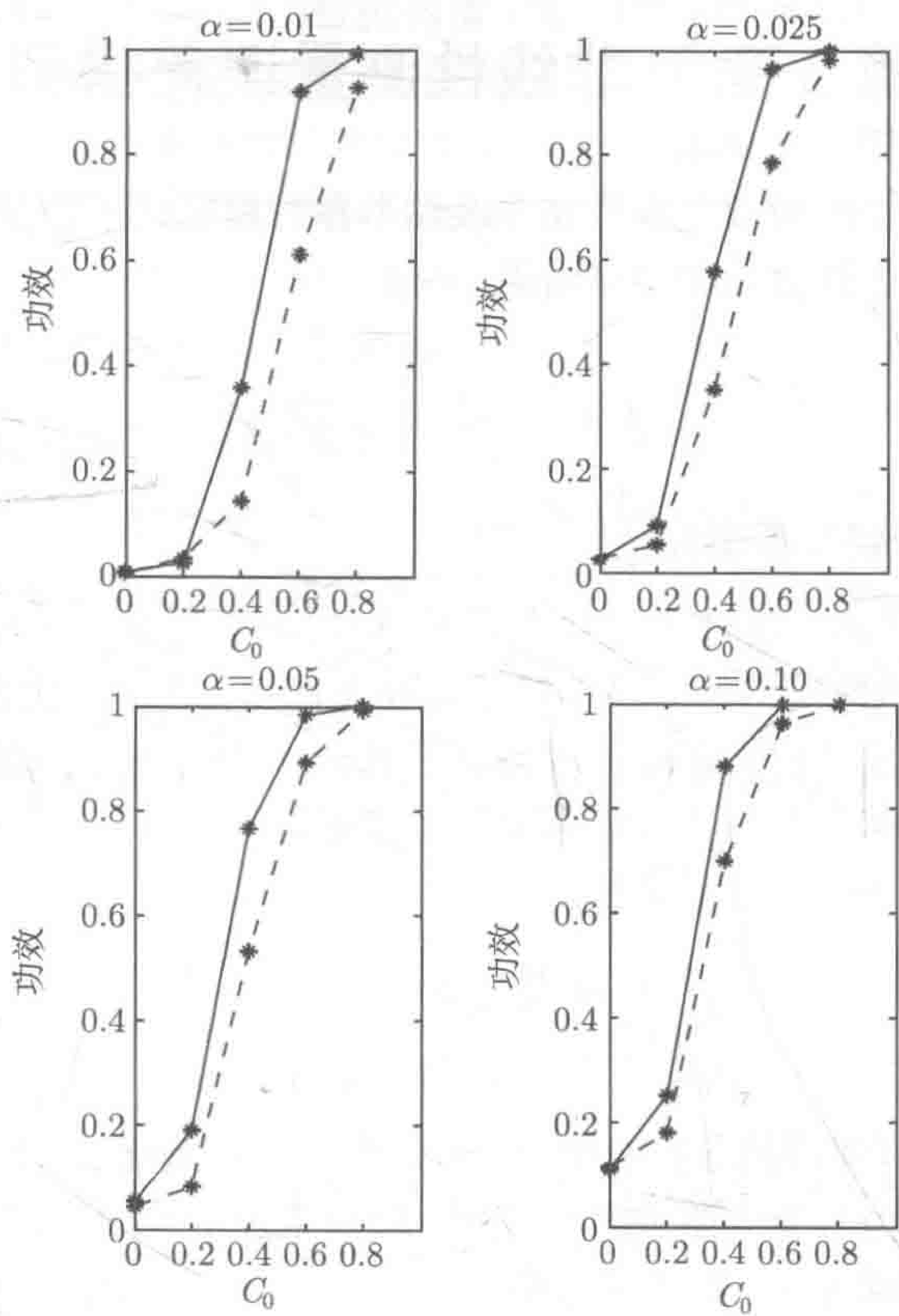


图 2.8.1 模型 (2.8.8) 的功效函数, 虚线表示 $n=300$ 的情形, 实线表示 $n=500$ 的情形

从图 2.8.1 可以看出, 在 $C_0=0$ 时, 基于 wild-bootstrap 重抽样得到的 p 值十分接近于预先选定的 4 个检验水平 0.01, 0.025, 0.05, 0.10. 而当 C_0 的取值变大时, wild-bootstrap 重抽样得到的 p 值越来越接近于 1. 表明 wild-bootstrap 重抽样方法是相合的, 并且样本量变大时, wild-bootstrap 重抽样方法效果变得越好.

第3章　非线性测量误差模型

本章考虑非线性测量误差模型和非线性半参数测量误差模型的统计推断问题, 介绍这两种模型的估计方法和理论结果.

3.1　引　　言

3.1.1　非线性测量误差模型介绍

现实世界中, 严格的线性模型并不多见, 同时随着科学技术和近代统计学的飞速发展, 不能化为线性模型的非线性模型也越来越多. 农业、生物、经济和工程技术等各个部门都提出了许多非线性模型以及其他非线性统计问题. 因此, 积极开展非线性模型估计的研究在理论和实践中日趋重要.

考虑如下非线性测量误差模型:

$$\begin{cases} Y = f(X, \beta) + \varepsilon, \\ W = X + u, \end{cases} \tag{3.1.1}$$

其中 X 是 q 维不可观测的随机变量, β 是 p 维未知参数向量. $f(\cdot,\cdot)$ 是 X 和 β 的已知可测函数, W 为 q 维可观测随机变量, 模型误差 ε 与 W 相互独立, 且均值为 0, 方差为 σ^2. 测量误差 u 与 X 相互独立, 且其均值为 0, 协方差矩阵为 Σ_{uu}. 考虑到模型的可识别性, 假设测量误差的协方差阵 Σ_{uu} 已知.

对模型 (3.1.1), 已经有许多统计学者在不同的假定下进行了研究, 得到了较理想的结果. 当 X 是固定未知常数时, Wolter 和 Fuller (1982) 考虑了误差变量趋于 0(样本容量 $n \to \infty$) 时的情况, 证明了参数的极大似然估计是相合的. 当 $f(\cdot,\cdot)$ 是多项式函数且误差方差已知时, Chan 和 Mak (1985) 证明了未知参数的最小二乘估计是 $\sqrt{n}$ 相合的. 对于 X 是随机变量的场合, Hsiao (1989) 证明了 β 的最小二乘估计是相合的且具有渐近正态分布. Li 等 (1996) 在假定 X, ε, u 都服从正态分布的情况下, 给出了 β 的估计量, 并证明了它是 β 的强一致相合估计并具有渐近正态性. Carroll 等 (1996) 研究了非线性 EV 模型 SIMEX 估计的渐近性质并构造了标准误差的估计. Buzas (1997) 基于工具变量方法研究了非线性 EV 模型的估计, 并证明了所得估计量的渐近性质. Taupin (2001) 利用改进的最小二乘估计准则提出了未知参数的一个相合估计 $\hat{\beta}$, 给出了估计量的收敛速度的一个上界, 并验证了对于一些特殊的回归函数, $\hat{\beta}$ 依概率收敛于 β 的速度可达到 $O(n^{-1/2})$. 进一步,

薛留根 (2005b) 研究了 $\hat{\beta}$ 的渐近性质, 证明了其渐近正态性, 同时也构造了模型误差方差的估计量, 并证明了所提估计量依概率收敛到 σ^2 的速度可达到 $O(n^{-1/2})$. Sepanskia 和 Lee (1995) 利用核实数据构造了非线性 EV 模型中未知参数的最小二乘估计, 并证明了所得估计量的相合性和渐近正态性. Stute 等 (2007) 把经验似然方法推广到具有核实数据的非线性 EV 模型, 构造了模型中未知参数的经验似然置信域. 更多关于非线性 EV 回归模型的详细讨论可参见文献 Lee 和 Sepanski (1995), Hsiao 等 (1997), Li (2002), Suh 和 Schafer (2002), Schennach (2004, 2007), Carroll 等 (2006), Liu 和 Wu (2008), Wang 和 Hsiao (2011) 等.

3.1.2 非线性半参数测量误差模型介绍

在许多实际问题中, 响应变量与某些协变量是通过回归参数函数相关而与其他变量的关系却不容易参数化. 在这种情形下, 人们自然会考虑利用半参数回归模型对数据进行拟合. 半参数回归模型保留了非参数模型的灵活性和参数模型容易解释的优点, 已经成为数据分析强有力的工具之一. Ruppert 等 (2003) 介绍了各种半参数模型以及它们的估计方法和应用, 其中一个重要的半参数模型是下面的非线性半参数回归模型:

$$Y = f(X, \beta) + g(T) + \varepsilon, \tag{3.1.2}$$

其中 Y 为实值响应变量, $f(\cdot,\cdot)$ 为已知可测函数, $g(\cdot)$ 是定义于闭区间 [0,1] 上的未知光滑函数. X 是 q 维随机变量, T 是取值于 [0,1] 上的随机变量, ε 是随机模型误差, $\beta \in \Theta$ 为 $p \times 1$ 未知参数向量. Li 和 Nie (2008) 研究了模型 (3.1.2) 中参数分量估计的相合性和渐近正态性等问题. 模型 (3.1.2) 是一类非常广泛的模型, 当 $f(X,\beta) = X^{\mathrm{T}}\beta$ 时, 该模型即为部分线性回归模型 (Härdle et al., 2000; 柴根象和洪圣岩, 1995); 当 $g(T) = 0$ 时, 该模型即为非线性参数回归模型 (Bates and Watts, 1988; Seber and Wild, 1989; Huet, 2004); 当 $f(X,\beta) = X^{\mathrm{T}}\beta$ 且 $g(T) = 0$ 时, 该模型即为线性回归模型; 当模型中仅仅存在非参数部分时, 模型即为非参数回归模型 (Härdle, 1990).

在实际问题中, 协变量 X 和 T 往往不能直接观测到或不能被准确地测出, 所能观测到的是受了误差干扰的变量 W 和 Z, 即

$$W = X + u, \tag{3.1.3}$$

$$Z = T + e. \tag{3.1.4}$$

对协变量 X 带有测量误差的情况已有许多文献讨论. 例如, 当 $f(X,\beta) = X^{\mathrm{T}}\beta$ 时, 部分线性 EV 模型作为非线性半参数 EV 模型的特殊情况, 已引起了统计学者的重视, 如 Liang 等 (1999) 在测量误差向量协方差阵已知的情形下, 获得了诸多参

数和非参数估计及其渐近性质, 他们提出的纠偏方法克服了由测量误差引起的参数估计的不相合性. Li 和 Xue (2008) 利用纠偏的经验似然构造了模型中兴趣参数的置信域. 崔恒建 (2004) 在 X 和 Y 有重复观测的情形下讨论了 β 的估计及其渐近性质, 并得到了非参数估计的最优收敛速度. Wang (1999) 与 Wang 和 Zhu (2001) 分别在核实数据和删失数据下研究了部分线性 EV 模型的估计问题, 并在一定条件下证明了估计量的渐近性质. 当 $f(X,\beta)$ 是 X 和 β 的非线性函数时, 崔恒建等 (1998) 在假定 X, u, ε 均服从正态分布的条件下证明了参数分量的强相合性和渐近正态性, 并给出了非参数分量的强一致收敛速度. 薛留根 (2006) 借助于核实数据, 利用经验似然方法构造了模型中参数 β 的置信域. 刘强 (2010) 在响应变量随机缺失情形下构造了非线性半参数 EV 模型中未知参数和非参数函数的两种估计, 并分别证明了所得估计的渐近正态性和最优收敛速度.

对协变量 T 带有测量误差的情形, 当 $f(X,\beta)=X^{\mathrm{T}}\beta$ 时, Liang (2000) 研究了模型的估计问题, 在适当条件下获得了参数和非参数函数的估计, 并证明了所得估计量的渐近性质. 进一步, Chen 和 Cui (2012) 与 Huang (2012) 分别构造了模型中未知参数的经验似然置信域. Zhu 和 Cui (2003) 在 X 和 T 都带有测量误差时, 利用矩方法得到了参数和非参数分量的估计及其渐近性质. 当 $f(X,\beta)$ 是 X 和 β 的非线性函数时, 冯三营等 (2010a) 与冯三营和薛留根 (2012) 研究了模型的经验似然推断问题, 构造了模型中未知参数的经验似然置信域及极大经验似然估计, 进一步也给出了非参数函数及模型误差方差的估计并证明了它们的渐近性质. Feng 等 (2014) 研究了模型的有效统计推断问题, 构造了模型中参数和非参数分量的估计并证明了它们的渐近性质.

3.2 非线性测量误差模型的估计方法

本节针对非线性测量误差模型 (3.1.1), 介绍四种估计方法：最小距离估计方法、调整的最小二乘估计方法、SIMEX 估计方法和工具变量估计方法.

3.2.1 最小距离估计方法

设 $\{(Y_i,W_i), i=1,\cdots,n\}$ 是来自模型 (3.1.1) 的一组 i.i.d. 随机样本. 假定 X 关于 W 的条件密度形式已知, 记为 $p(x|W;\delta)$, 其中 δ 为 $d\times 1$ 维未知参数向量. 由 ε 和 W 相互独立, 可得给定 W_i 条件下 Y_i 的条件均值为

$$E(Y_i|W_i)=\int f(x_i,\beta)p(x_i|W_i;\delta)\mathrm{d}x_i =: G(W_i;\theta), \tag{3.2.1}$$

其中 $\theta=(\beta^{\mathrm{T}},\delta^{\mathrm{T}})^{\mathrm{T}}$ 为 $(p+d)\times 1$ 维参数向量. 因此, 模型 (3.1.1) 中的第一个式子

可以重写为

$$Y_i = G(W_i;\theta) + \epsilon_i, \tag{3.2.2}$$

其中 $\epsilon_i = \varepsilon_i + f(X_i,\beta) - G(W_i;\theta)$, 且有 $E\epsilon_i = 0$,

$$\sigma_i^2 = \mathrm{Var}(\epsilon_i|W_i) = \sigma^2 + \int f(x,\beta)^2 p(x_i|W_i;\delta)\mathrm{d}x_i - G(W_i;\theta)^2.$$

Hsiao (1989) 研究了模型 (3.2.2) 的可识别性问题, 并提出一个非线性最小距离估计方法, 即最小化目标函数

$$S_n = \frac{1}{n}\sum_{i=1}^{n}(Y_i - G(W_i;\theta))^2, \tag{3.2.3}$$

可得未知参数 θ 的估计 $\hat{\theta}$. 在一定正则条件下, Hsiao (1989) 证明了 $\hat{\theta}$ 的相合性和渐近正态性.

由于在利用上述方法估计 θ 时, 需要计算多重积分 $\int f(x,\beta)p(x|W;\delta)\mathrm{d}x$, 尤其是当 X 的维数较高时, 这个计算是相当困难的. 为了避免计算多重积分, Li (2000) 提出了一个基于模拟的最小距离估计方法. 具体思想如下, 注意到

$$\begin{aligned} G(W_i;\theta) &= \int f(x_i,\beta)p(x_i|W_i;\delta)\mathrm{d}x_i \\ &= \int \frac{f(x_i,\beta)p(x_i|W_i;\delta)r_i(x_i)}{r_i(x_i)}\mathrm{d}x_i \\ &= E_{r_i}\left[\frac{f(x_i,\beta)p(x_i|W_i;\delta)}{r_i(x_i)}\right], \end{aligned} \tag{3.2.4}$$

其中 $r_i(\cdot)$ 为已知 q 维密度函数, $E_{r_i}(\cdot)$ 表示关于 $r_i(\cdot)$ 的期望.

对每一个 $i=1,\cdots,n$, 由 $r_i(\cdot)$ 产生独立同分布随机样本 $\{X_{hi}, h=1,\cdots,H\}$, 则由式 (3.2.4), $G(W_i;\theta)$ 可以近似为 $\bar{G}_i(\theta) = 1\Big/H\sum_{j=1}^{H} G_{hi}(\theta)$, 其中

$$G_{hi}(\theta) = \frac{f(X_{hi},\beta)p(X_{hi}|W_i;\delta)}{r_i(X_{hi})}.$$

于是式 (3.2.3) 转化为

$$S_{H,n} = \frac{1}{n}\sum_{i=1}^{n}\left[(Y_i - \bar{G}_i(\theta))^2 - \frac{1}{H(H-1)}\sum_{h=1}^{H}(G_{hi}(\theta) - \bar{G}_i(\theta))^2\right]. \tag{3.2.5}$$

最小化 $S_{H,n}$ 可得 θ 的基于模拟的最小距离估计 $\hat{\theta}_S$. Li (2000) 在适当条件下证明了 $\hat{\theta}_S$ 是 θ 的强相合估计, 并得到了 $\hat{\theta}_S$ 的渐近正态分布.

3.2.2 调整的最小二乘估计

假定模型 (3.1.1) 中的 ε 和 u 都服从均值为 0 的正态分布, 且 $\{(Y_i, W_i), i = 1, \cdots, n\}$ 是来自模型 (3.1.1) 的一组 i.i.d. 随机样本. 若协变量 X 的概率密度函数 $g(\cdot)$ 已知, 则一个自然的估计准则定义为

$$\tilde{S}_n(\beta, g) = \frac{1}{n}\sum_{i=1}^{n}\Psi(W_i)[Y_i - E(f(X_i, \beta)|W_i)]^2, \tag{3.2.6}$$

其中 $\Psi(\cdot)$ 是一个具有紧支撑的权函数, β 的估计定义为

$$\tilde{\beta} = \arg\min_{\beta} \tilde{S}_n(\beta, g).$$

可以证明 $\tilde{\beta}$ 是 β 的一个相合估计且具有渐近正态性. 然而在模型 (3.1.1) 中, $g(\cdot)$ 是未知的, 条件期望 $E(f(X_i, \beta)|W_i)$ 不仅依赖于 W_i 也依赖于未知密度 $g(\cdot)$. 因而 $\tilde{\beta}$ 显然不是一个估计. 为此, Taupin (2001) 采用基于样本 $W_1, \cdots, W_n$ 的非参数估计来替代 $\tilde{S}_n(\beta, g)$ 中的未知条件期望 $E(f(X_i, \beta)|W_i)$, 从而定义如下调整的最小二乘准则

$$S_n(\beta) = \frac{1}{n}\sum_{i=1}^{n}\Psi(W_i)[Y_i - \hat{\Phi}_{f_\beta}(W_i)]^2, \tag{3.2.7}$$

其中 $\hat{\Phi}_{f_\beta}(W_i)$ 是 $E(f(X_i, \beta)|W_i)$ 的基于样本 $W_1, \cdots, W_n$ 的估计. 由此可以定义 β 的估计

$$\hat{\beta} = \arg\min_{\beta} S_n(\beta). \tag{3.2.8}$$

下面给出 $\hat{\Phi}_{f_\beta}(\cdot)$ 的具体形式. 首先引入一些记号, 用 u^* 表示 $u(x)$ 的 Fourier 变换, $\|u\|_1 = \int |u(x)|\mathrm{d}x$ 表示函数 $u(x)$ 的 $L_1(R)$-范数, $\langle u, v\rangle = \int u(x)\bar{v}(x)\mathrm{d}x$ 表示 $L_2(R)$ 中的内积. 引入函数 $I_C(t) = I_{|t|\leqslant C}$, $\tilde{I}_C(t) = I_{|t|\geqslant C}$, $T_{f_\beta, w}(x) = f(x, \beta)\eta(w - x)$, $T^*_{f_\beta, w}(t) = \int T_{f_\beta, w}(x)e^{-\mathrm{i}tx}\mathrm{d}x$, 这里 $\eta(\cdot)$ 表示标准正态密度. 记 $\|u\|_{\infty, \Psi} = \sup_{x\in\Delta}\|u(x)\|$, 其中 Δ 是权函数 $\Psi(\cdot)$ 的支撑集. 由 X 与 u 的独立性可得

$$E(f(X, \beta)|W) = \frac{\int f(x, \beta)\eta(W - x)g(x)\mathrm{d}x}{\int \eta(W - x)g(x)\mathrm{d}x} =: \frac{\Gamma_{f_\beta}(W)}{h(W)} =: \Phi_{f_\beta}(W), \tag{3.2.9}$$

其中 $h = \eta * g$($*$ 表示卷积) 是 W 的密度.

由式 (3.2.9), 首先估计 $\Gamma_{f_\beta}(w)$. 利用 X 与 u 的独立性及 Parseval-Plancherel 公式可得

$$\Gamma_{f_\beta}(w) = (2\pi)^{-1}\langle T^*_{f_\beta,w}, g^*\rangle = (2\pi)^{-1}\langle T^*_{f_\beta,w}(\eta^*)^{-1}, h^*\rangle.$$

$h^*(t)$ 的经验估计为

$$\hat{h}^*(t) = \frac{1}{n}\sum_{j=1}^{n} e^{-\mathrm{i}tW_j} K_n^*(t) =: \frac{1}{n}\sum_{j=1}^{n} h^*_{n,j}(t),$$

这里 $K_n^*(t)$ 是函数 $K_n(x) = C_n K(C_n x)$ 的 Fourier 变换, K 是待定的核函数, C_n 是趋于 ∞ 的正数序列. 于是, 定义 $\Gamma_{f_\beta}(w)$ 的估计为

$$\hat{\Gamma}_{f_\beta}(w) = \frac{1}{2\pi n}\sum_{j=1}^{n} \mathrm{Re}\langle T^*_{f_\beta,w}(\eta^*)^{-1}, h^*_{n,j}\rangle, \tag{3.2.10}$$

其中 $\mathrm{Re}(z)$ 表示 z 的实部.

进一步, 由于分母 $h(w)$ 为密度函数, 因此可以采用核估计的方法得到. 此外, 注意到 $h(w) = \Gamma_1(w)$, 因此可以选取与式 (3.2.10) 中相同的核函数来估计 $h(w)$, 即定义 $h(w)$ 的估计为

$$\hat{h}(w) = \frac{1}{n}\sum_{j=1}^{n} C_n^{-1} K(C_n^{-1}(w - W_j)),$$

从而可得 $\Phi_{f_\beta}(w)$ 的估计为

$$\hat{\Phi}_{f_\beta}(w) = \hat{\Gamma}_{f_\beta}(w)/\hat{h}(w). \tag{3.2.11}$$

关于权函数 $\Psi(\cdot)$ 和核函数 $K(\cdot)$ 的选取可以见 Taupin (2001). Taupin (2001) 证明了 $\hat{\beta}$ 是 β 的一个相合估计, 并给出了 $\hat{\beta}$ 的收敛速度的一个上界. 薛留根 (2005b) 进一步研究了 $\hat{\beta}$ 的渐近性质, 证明了 $\hat{\beta}$ 具有渐近正态性.

3.2.3 SIMEX 估计方法

设 $\{(Y_i, W_i),\ i = 1, \cdots, n\}$ 是来自模型 (3.1.1) 的一组 i.i.d.随机样本, 并假设 Σ_{uu} 已知. 当 Σ_{uu} 未知时, 可采用 2.2 节中介绍的重复测量的方法对其进行估计得到. 若 X_i 能够准确观测, 则可以通过求解如下估计方程得到未知参数 β 的相合估计

$$0 = \sum_{i=1}^{n} \frac{\partial f(X_i,\beta)}{\partial\beta}(Y_i - f(X_i,\beta)) =: \sum_{i=1}^{n}\psi(Y_i, X_i; \beta). \tag{3.2.12}$$

由于存在测量误差, 若直接用 W_i 替代式 (3.2.12) 中的 X_i, 则所得 β 的估计将不再是相合估计. SIMEX 是基于模拟的方法来估计未知参数并减少测量误差带来的估计偏差. 下面介绍 SIMEX 估计程序.

模拟步 (simulation step): 对于每一个 $i = 1, \cdots, n$, 产生一系列变量

$$W_{ib}(\lambda) = W_i + (\lambda \Sigma_{uu})^{1/2} U_{ib}, \quad b = 1, \cdots, B,$$

其中 U_{ib} 是随机产生的服从 q 维多元正态分布 $N(0, I_q)$ 的独立随机向量, I_q 是 $q \times q$ 的单位矩阵, B 是一个给定的任意大的正整数, $\lambda \in \Lambda = \{\lambda_1, \lambda_2, \cdots, \lambda_M\}$, 这里 Λ 表示 λ 取值的格子点.

估计步 (estimation step): 对每一个 b, 记 $\hat{\beta}_b(\lambda)$ 为下式的解

$$0 = \sum_{i=1}^{n} \psi(Y_i, W_{ib}; \beta). \tag{3.2.13}$$

下面, 对所得估计 $\hat{\beta}_1(\lambda), \cdots, \hat{\beta}_B(\lambda)$ 进行平均, 可得

$$\hat{\beta}(\lambda) = \frac{1}{B} \sum_{b=1}^{B} \hat{\beta}_b(\lambda).$$

若取 $\lambda = 0$, 则 $\hat{\beta}(0)$ 表示忽略测量误差后得到的 naive 估计, 即直接用 W_i 代替 X_i 得到的最小二乘估计.

外推步 (extrapolation step): 记 $\mathcal{G}(\lambda, \Gamma)$ 为一个外推函数, 其中 Γ 为未知参数向量. 在 $\lambda \in \Lambda$ 上, 以外推函数 $\mathcal{G}(\lambda, \Gamma)$ 来拟合 $\hat{\beta}(\lambda)$, 从而可得 Γ 的估计 $\hat{\Gamma}$. 最终, β 的 SIMEX 估计定义为

$$\hat{\beta}_{\text{SIMEX}} = \mathcal{G}(-1, \hat{\Gamma}). \tag{3.2.14}$$

Carroll 等 (1996) 证明了 $\hat{\beta}_{\text{SIMEX}}$ 的渐近性质. 在实际运算中, 关于外推函数 $\mathcal{G}(\lambda, \Gamma)$ 的选取可参见文献 Cook 和 Stefanski (1994), Carroll 等 (1996), Liang 和 Ren (2005), Lin 和 Carroll (2000) 等, 和本书 1.6.3 节.

3.2.4 工具变量方法

设 $\{(Y_i, W_i, T_i), i = 1, \cdots, n\}$ 是来自模型 (3.1.1) 的一组 i.i.d. 随机样本, 其中 T_i 为工具变量. 若 X_i 能够准确观测, 则可以通过求解如下估计方程得到来估计未知参数 β,

$$\sum_{i=1}^{n} \psi(Y_i, X_i; \beta, \tau) = 0, \tag{3.2.15}$$

其中 $\psi(Y,X;\beta,\tau)=\{Y-f(X,\beta)\}g(X;\beta,\tau)$, $g(X;\beta,\tau)$ 为 $p\times 1$ 维函数, τ 为已知模型散度参数.

上述方法涵盖了一类常见估计方法. 例如, 若取 $g(X;\beta,\tau)=\dfrac{\partial}{\partial\beta}f(X,\beta)$ (Gallant, 1987), 则对应于具有常数方差的非线性回归模型的最小二乘估计; 若取

$$g(X;\beta,\tau)=\left[\frac{\partial}{\partial\beta}f(X,\beta)\right]\Big/V(X;\beta,\tau),$$

其中 $V(X;\beta,\tau)$ 为响应变量 Y 的方差函数, 则对应于拟似然方法 (Carroll and Ruppert, 1988).

由于存在测量误差, 若直接用 W_i 替代式 (3.2.15) 中的 X_i, 则所得 β 的估计将不再是相合估计. Buzas (1997) 借助于工具变量 T_i, 给出了一个新的模型估计方法. 其基本思想为, 若在给定 (X,T) 的条件下 u 的分布关于 0 点对称, 则定义

$$\phi(W,T;\beta)=\left|\frac{f_x(E(X|T),\beta)}{f_x(W,\beta)}\right|,$$

其中 $f_x(x,\beta)=\partial f(x,\beta)/\partial x$. 从而构造调整的得分函数

$$\psi_{\mathrm{IV}}(Y_i,W_i,T_i;\beta,\tau)=\phi(W_i,T_i;\beta)[Y_i-f(W_i,\beta)]g(E(X_i|T_i);\beta,\tau). \tag{3.2.16}$$

于是, β 的工具变量估计 $\hat\beta_{\mathrm{IV}}$ 可以通过求解如下估计方程得到

$$\sum_{i=1}^{n}\psi_{\mathrm{IV}}(Y_i,W_i,T_i;\beta,\tau)=0. \tag{3.2.17}$$

Buzas (1997) 证明了 $\hat\beta_{\mathrm{IV}}$ 是 β 的一个相合估计, 并证明了 $\hat\beta_{\mathrm{IV}}$ 的渐近正态性.

3.3 非线性测量误差模型的经验似然推断

本节考虑非线性测量误差模型 (3.1.1), 构造模型中未知参数的经验对数似然比统计量. 此外, 为了克服计算困难, 提出了一个基于模拟的经验对数似然比统计量. 在适当条件下, 证明了所提出的两种统计量都具有渐近 χ^2 分布, 所得结果可以用来构造未知参数的置信域.

3.3.1 经验似然方法

设 $(W_1,Y_1),\cdots,(W_n,Y_n)$ 是来自模型 (3.1.1) 的一组独立同分布的随机样本, 即

$$\begin{cases} Y_j=f(X_j,\beta)+\varepsilon_j,\\ W_j=X_j+u_j, \quad j=1,2,\cdots,n, \end{cases} \tag{3.3.1}$$

其中 $(\varepsilon_1, u_1), \cdots, (\varepsilon_n, u_n)$ 为 i.i.d.均值为 0 的向量, 且 $\{\varepsilon_j, 1 \leqslant j \leqslant n\}$ 与 $\{W_j, 1 \leqslant j \leqslant n\}$ 相互独立, $\{u_j, 1 \leqslant j \leqslant n\}$ 与 $\{W_j, 1 \leqslant j \leqslant n\}$ 相互独立, $\mathrm{Var}(\varepsilon_1) = \sigma^2$, $0 < \sigma^2 < \infty$, $\mathrm{Cov}(u_1) = \Sigma_{uu}$.

假设 X 关于 W 的条件分布的形式已知, 记为 $g(x|W;\delta)$, 其中 δ 为 $m \times 1$ 未知向量, 则

$$E(Y_j|W_j) = \int f(x_j, \beta) g(x_j|W_j;\delta) \mathrm{d}x_j =: G(W_j;\gamma), \tag{3.3.2}$$

其中 $\gamma = (\beta^{\mathrm{T}}, \delta^{\mathrm{T}})^{\mathrm{T}}$ 为 $(p+m) \times 1$ 常向量. 用 $G(W_j;\gamma)$ 替代式 (3.3.1) 中的 $f(X_j, \beta)$, 可得

$$Y_j = G(W_j;\gamma) + \omega_j, \tag{3.3.3}$$

其中 $\omega_j = \varepsilon_j + f(X_j, \beta) - G(W_j;\gamma)$, 易知 $E\omega_j = 0$.

定义辅助随机向量

$$Z_j(\gamma) = [Y_j - G(W_j;\gamma)] \frac{\partial G(W_j;\gamma)}{\partial \gamma}. \tag{3.3.4}$$

当 γ 是真实参数时, $E(Z_j(\gamma)) = 0$. 因此, 定义经验对数似然比函数为

$$\mathcal{R}(\gamma) = -2 \max \left\{ \sum_{i=1}^{n} \log(np_i) \,\middle|\, p_i \geqslant 0,\ \sum_{i=1}^{n} p_i = 1,\ \sum_{i=1}^{n} p_i Z_j(\gamma) = 0 \right\}, \tag{3.3.5}$$

由 Lagrange 乘子法, $\mathcal{R}(\gamma)$ 可以表示为

$$\mathcal{R}(\gamma) = 2 \sum_{j=1}^{n} \log\{1 + \lambda^{\mathrm{T}} Z_j(\gamma)\}, \tag{3.3.6}$$

其中 λ 是 Lagrange 乘子, 满足

$$\frac{1}{n} \sum_{j=1}^{n} \frac{Z_j(\gamma)}{1 + \lambda^{\mathrm{T}} Z_j(\gamma)} = 0. \tag{3.3.7}$$

为了得到 $\mathcal{R}(\gamma)$ 的渐近分布, 需要下面的条件.

(C3.3.1) 参数空间 $\Gamma = \Theta \times \Delta \subset \mathbb{R}^{p+m}$ 是凸紧空间.

(C3.3.2) 对于任意的 $\beta \in \Theta$, $f(x, \beta)$ 是 x 的可测函数, 并且对于任意的 x, $f(x, \beta)$ 关于 β 连续.

(C3.3.3) $f(x, \beta)$, $g(x_j|W_j;\delta)$ 分别对 β 和 δ 有连续的二阶导数.

(C3.3.4) $\int \sup\limits_{\Theta \times \Delta} f^2(x, \beta) g(x|W;\delta) \mathrm{d}x < \infty$.

(C3.3.5) 存在可测函数 $L(x)$, $F(x)$ 使得在真参数 γ 的邻域 $N\subset\Gamma$ 内, 有下式成立:

(1) $\left|\dfrac{\partial G(x;\gamma)}{\partial\gamma}\right|\leqslant L(x)$,

(2) $\left|\dfrac{\partial G^2(x;\gamma)}{\partial\gamma\partial\gamma^{\mathrm{T}}}\right|\leqslant F(x)$,

(3) $E(L(x))<\infty$, $E[L(x)L(x)^{\mathrm{T}}]<\infty$.

(C3.3.6) 矩阵 C, C_s 是正定矩阵, 其中

$$C=E\left[\omega_1^2\frac{\partial G(W_1;\gamma)}{\partial\gamma}\frac{\partial G(W_1;\gamma)}{\partial\gamma^{\mathrm{T}}}\right],\quad C_s=\mathrm{Var}(\widetilde{Z}_1(\gamma)).$$

定理 3.3.1 假设条件(C3.3.1)~条件(C3.3.6)成立, 如果 γ 是参数真值, 当 $n\to\infty$ 时, 则有 $\mathcal{R}(\gamma)\xrightarrow{L}\chi^2_{p+m}$.

基于定理 3.3.1, 可以定义参数向量 γ 的置信域, 即对任给的 $0<\alpha<1$, 存在 c_α 使得 $P\{\chi^2_{p+m}\leqslant c_\alpha\}=1-\alpha$, 则

$$I_\alpha(\gamma^*)=\left\{\gamma^*\ \middle|\ \mathcal{R}(\gamma^*)\leqslant c_\alpha\right\}$$

为参数向量 γ 的具有渐近置信水平 $1-\alpha$ 的置信域, 并且有

$$P(\gamma\in I_\alpha(\gamma^*))=1-\alpha+o(1).$$

由于添加了 m 维未知参数 δ, 使得所构造的经验对数似然比统计量依分布渐近收敛到自由度为 $p+m$ 的标准 χ^2 分布, 而通常的研究对象主要是 p 维的参数向量 β. 由于自由度的增加, 使得由定理 3.3.1 得到的 β 的置信域精度会降低. 若假设 $u\sim N(0,1)$, X 的密度函数为 $g(x)$, 则式 (3.3.2) 转化为

$$\begin{aligned}E(Y_j|W_j)&=E(f(X_j,\beta)|W_j)\\&=\frac{\displaystyle\int f(x_j,\beta)\phi(W_j-x_j)g(x_j)\mathrm{d}x_j}{\displaystyle\int\phi(W_j-x_j)g(x_j)\mathrm{d}x_j}=:Q(W_j;\beta),\end{aligned}\tag{3.3.8}$$

其中 $\phi(\cdot)$ 是标准正态密度函数.

记 $Z_j(\beta)=[Y_j-Q(W_j;\beta)]\dfrac{\partial Q(W_j;\beta)}{\partial\beta}$, 则式 (3.3.6) 改写为

$$\mathcal{R}(\beta)=2\sum_{j=1}^{n}\log\{1+\lambda^{\mathrm{T}}Z_j(\beta)\}.\tag{3.3.9}$$

定理 3.3.2　假设条件(C3.3.1)~条件(C3.3.6)成立, 如果 β 是参数真值, 当 $n \to \infty$ 时, 则有 $\mathcal{R}(\beta) \xrightarrow{L} \chi_p^2$.

基于定理 3.3.2, 可以构造参数向量 β 的置信域, 即对任给的 $0 < \alpha < 1$, 存在 c_α 使得 $P\{\chi_p^2 \leqslant c_\alpha\} = 1 - \alpha$, 则

$$I_\alpha(\beta^*) = \left\{\beta^* \mid \mathcal{R}(\beta^*) \leqslant c_\alpha\right\}$$

为参数向量 β 的具有渐近置信水平 $1-\alpha$ 的置信域, 并且有

$$P(\beta \in I_\alpha(\beta^*)) = 1 - \alpha + o(1).$$

3.3.2　基于模拟的经验似然方法

由于构造经验似然统计量 $\mathcal{R}(\gamma)$ 时需要计算多重积分 $\int f(x,\beta)g(x|W;\delta)\mathrm{d}x$, 尤其是当协变量 X 的维数高于 3 维时, 这个计算是相当困难的. 为了避免计算多重积分, Li (2000) 和 Wang (2004) 分别使用基于模拟的方法来处理这个问题. 受此启发, 本节提出一个基于 Monte Carlo 模拟的经验似然方法.

由式 (3.3.2)

$$\begin{aligned} G(W_j;\gamma) &= \int f(x_j,\beta)g(x_j|W_j;\delta)\mathrm{d}x_j \\ &= \int \frac{f(x_j,\beta)g(x_j|W_j;\delta)\varphi(x_j)}{\varphi(x_j)}\mathrm{d}x_j \\ &= E_\varphi\left[\frac{f(X_j,\beta)g(X_j|W_j;\delta)}{\varphi(X_j)}\right], \end{aligned} \tag{3.3.10}$$

其中 $\varphi(\cdot)$ 是一个已知密度函数, $E_\varphi[\cdot]$ 表示关于 $\varphi(\cdot)$ 的期望.

由 $\varphi(x)$, 对每一个 j $(j = 1, \cdots, n)$, 产生 i.i.d.随机样本 $\{X_{js}, s = 1, 2, \cdots, 2S\}$. 记

$$G_S(W_j;\gamma) = \frac{1}{S}\sum_{s=1}^{S}\frac{f(X_{js},\beta)g(X_{js}|W_j;\delta)}{\varphi(X_{js})},$$

$$G_{2S}(W_j;\gamma) = \frac{1}{S}\sum_{s=S+1}^{2S}\frac{f(X_{js},\beta)g(X_{js}|W_j;\delta)}{\varphi(X_{js})},$$

则依概率有

$$G_S(W_j;\gamma) \longrightarrow G(W_j;\gamma),$$

$$G_{2S}(W_j;\gamma) \longrightarrow G(W_j;\gamma).$$

定义辅助随机变量

$$\begin{aligned}\widetilde{Z}_j(\gamma) =& \frac{1}{2}[Y_j - G_S(W_j;\gamma)]\frac{\partial G_{2S}(W_j;\gamma)}{\partial\gamma}\\ &+\frac{1}{2}[Y_j - G_{2S}(W_j;\gamma)]\frac{\partial G_S(W_j;\gamma)}{\partial\gamma}.\end{aligned} \tag{3.3.11}$$

当 γ 是真实参数时, $E(\widetilde{Z}_j(\gamma)) = E[E(\widetilde{Z}_j(\gamma)|W_j)] = 0$. 从而, 定义基于模拟的经验对数似然比函数为

$$\widetilde{\mathcal{R}}(\gamma) = -2\max\left\{\sum_{i=1}^{n}\log(np_i)\;\middle|\; p_i \geqslant 0,\ \sum_{i=1}^{n}p_i = 1,\ \sum_{i=1}^{n}p_i\widetilde{Z}_j(\gamma) = 0\right\}. \tag{3.3.12}$$

由 Lagrange 乘子法, $\widetilde{\mathcal{R}}(\gamma)$ 可以表示为

$$\widetilde{\mathcal{R}}(\gamma) = 2\sum_{j=1}^{n}\log\{1+\widetilde{\lambda}^{\mathrm{T}}\widetilde{Z}_j(\gamma)\}, \tag{3.3.13}$$

其中 $\widetilde{\lambda}$ 满足

$$\frac{1}{n}\sum_{j=1}^{n}\frac{\widetilde{Z}_j(\gamma)}{1+\widetilde{\lambda}^{\mathrm{T}}\widetilde{Z}_j(\gamma)} = 0.$$

定理 3.3.3 假设 $\varphi(x)$ 的支撑包含 $g(x|W;\delta)$ 的支撑, 且条件(C3.3.1)~条件(C3.3.6)成立, 如果 γ 是参数真值, 当 $n\to\infty$ 时, 则有 $\widetilde{\mathcal{R}}(\gamma)\xrightarrow{L}\chi^2_{p+m}$.

基于定理 3.3.3, 可以定义参数向量 γ 的置信域, 即对任给的 $0<\alpha<1$, 存在 c_α 使得 $P\{\chi^2_{p+m}\leqslant c_\alpha\} = 1-\alpha$, 则

$$\widetilde{I}_\alpha(\gamma^*) = \left\{\gamma^* \mid \widetilde{\mathcal{R}}(\gamma^*)\leqslant c_\alpha\right\}$$

为参数向量 γ 的具有渐近置信水平 $1-\alpha$ 的置信域, 并且有 $P(\gamma\in\tilde{I}_\alpha(\gamma^*)) = 1-\alpha+o(1)$.

推论 3.3.1 设定理 3.3.3 的条件成立, 如果 γ 是参数真值, 则有

$$\begin{aligned}C_S =& C+\frac{1}{S}E\left[\frac{\partial(Y_1-G_1)(G_1-G_{11})}{\partial\gamma}\frac{\partial(G_1-G_{11})(Y_1-G_1)}{\partial\gamma^{\mathrm{T}}}\right]\\ &+\frac{1}{S^2}E\left[\frac{\partial(G_1-G_{11})(G_1-G_{12})}{\partial\gamma}\frac{\partial(G_1-G_{12})(G_1-G_{11})}{\partial\gamma^{\mathrm{T}}}\right],\end{aligned}$$

其中 $G_j = G(W_j;\gamma)$, $G_{js} = \dfrac{f(X_{js},\beta)g(X_{js}|W_j;\delta)}{\varphi(X_{js})}$.

一个值得注意的问题是密度函数 $\varphi(x)$ 与模拟样本数 S 如何选取. 首先, 从定理 3.3.3 和推论 3.3.1 可以看出, S 越大得到的参数置信域精度越高. 与定理 3.3.1 相比较, 基于模拟方法得到的参数的置信域的有效性损失率为 $O(1/S)$. 其次, 理论上 $\varphi(x)$ 的选取只要满足这个密度函数的支撑包含 $g(x|W;\delta)$ 的支撑即可. 但是在实际应用中, $\varphi(x)$ 一般应选取与被积函数 $f(x,\beta)g(x|W;\delta)$ 尽可能相似的密度函数, 使得 $f(x,\beta)g(x|W;\delta)/\varphi(x)$ 接近于常数, 从而提高结果的精度.

定理 3.3.1~定理 3.3.3 和推论 3.3.1 的详细证明可参考冯三营等 (2007).

3.3.3 模拟研究

本节通过模拟研究比较所提两种方法得到的未知参数 β 的置信域的精度. 为简单起见, 这里假定真实协变量 $X \sim N(1,1)$, 测量误差 $u \sim N(0,1)$.

考虑模型:

$$Y = 1.2(\beta X - 1)^2 + \varepsilon, \quad W = X + u,$$

其中 $\beta = 1$, $\varepsilon \sim N(0, 0.4^2)$. 选取密度函数 $\varphi(\cdot)$ 分别为服从自由度为 1 的 t 分布和均值为 2 方差为 1 的正态分布的密度函数. 取样本容量分别为 n=50, 100, 200, 随机模拟样本容量 $2S$=60, 120, 240, 480. 简记经验似然方法为 EL, 基于正态分布模拟的经验似然方法为 NEL, 基于 t 分布模拟的经验似然方法为 TEL, 重复模拟次数 M=1000 次, 模拟结果在表 3.3.1 中给出.

表 3.3.1 β 的置信度为 95%的置信区间的经验覆盖概率

n	EL	2S	NEL	TEL
$n = 50$	0.9140	60	0.8860	0.8790
		120	0.8940	0.8850
		240	0.9070	0.9020
		480	0.9110	0.9090
$n = 100$	0.9310	60	0.9030	0.8940
		120	0.9120	0.9050
		240	0.9210	0.9130
		480	0.9290	0.9260
$n = 200$	0.9420	60	0.9080	0.8980
		120	0.9220	0.9190
		240	0.9350	0.9270
		480	0.9410	0.9350

从模拟结果可以看出, 随着样本容量 n 的增大, 覆盖概率趋于置信度 0.95. 对比经验似然方法和基于模拟的经验似然方法, 从模拟过程可以知道, 基于模拟的经验似然方法的计算速度远远快于应用积分算法的经验似然方法; 同时从模拟结果也可以看出随着模拟样本个数 $2S$ 的增加, 基于模拟方法所得的覆盖概率逐渐趋于真

实结果. 另外, 对比 NEL 方法和 TEL 方法可以看出, 当 $\varphi(x)$ 的选取与被积函数 $f(x,\beta)$ $g(x|W;\delta)$ 更接近时, 所得结果的精度更高. 这也进一步说明了前面所叙述的结论的正确性.

3.4 非线性半参数 EV 模型的 profile 最小二乘估计

本节考虑非参数协变量带有测量误差的非线性半参数模型, 结合逆卷积的思想, 提出纠偏的 profile 非线性最小二乘方法, 获得模型中参数分量和非参数分量的估计, 在一定条件下, 证明参数分量估计的相合性和渐近正态性, 以及非参数分量估计的收敛速度.

3.4.1 估计方法

假设记录数据 $\{X_i, Z_i, Y_i\}_{i=1}^n$ 是来自模型 (3.1.2) 和 (3.1.4) 的一组 i.i.d. 的可观测随机样本, 即

$$\begin{cases} Y_i = f(X_i,\beta) + g(T_i) + \varepsilon_i, \\ Z_i = T_i + e_i, \end{cases} \tag{3.4.1}$$

其中 $\{\varepsilon_i\}_{i=1}^n$ 是相互独立的模型误差, 且 $E(\varepsilon_i) = 0$, $\operatorname{Var}(\varepsilon_i) = \sigma^2 < \infty, i = 1, 2, \cdots, n$. 测量误差 $\{e_i\}_{i=1}^n$ 独立同分布, 服从均值为 0, 方差为 σ_e^2 的已知分布. e_i 与 $(X_i, T_i, \varepsilon_i)$ 相互独立, ε_i 与 (X_i, T_i) 相互独立.

对于给定的 β, 模型 (3.1.2) 可以写为下面的非参数回归模型:

$$Y - f(X,\beta) = g(T) + \varepsilon.$$

因此若 β 已知, 当 T 可准确观测时, 由 $g(T) = E(Y - f(X,\beta)|T)$, 则 $g(\cdot)$ 能够采用下面的一些光滑方法进行估计, 如核方法、局部多项式方法、正交序列逼近方法、光滑样条方法、B 样条方法和小波估计方法等. 为简单起见, 这里采用核方法, 即定义 $g(\cdot)$ 的核估计为

$$\begin{aligned} \hat{g}(t;\beta) &= \sum_{i=1}^n K\left(\frac{t-T_i}{h}\right)(Y_i - f(X_i,\beta)) \Big/ \sum_{i=1}^n K\left(\frac{t-T_i}{h}\right) \\ &=: \sum_{i=1}^n W_i(t)(Y_i - f(X_i,\beta)), \end{aligned} \tag{3.4.2}$$

其中

$$W_i(\cdot) = \frac{1}{nh} K\left(\frac{\cdot - T_i}{h}\right) \Big/ \hat{r}(\cdot), \tag{3.4.3}$$

$K(\cdot)$ 为已知核函数, h 为光滑参数, $\hat{r}(t) = \dfrac{1}{nh}\sum_{i=1}^{n} K\left(\dfrac{t-T_i}{h}\right)$ 是协变量 T 的密度函数 $r(t)$ 的核估计. 从而可以定义如下的 profile 非线性最小二乘目标函数:

$$\sum_{i=1}^{n}\{Y_i - \hat{g}(T_i;\beta) - f(X_i,\beta)\}^2. \tag{3.4.4}$$

当 T_i 能够准确测量时, 极小化 profile 非线性最小二乘目标函数 (3.4.4), 可以得到参数分量 β 的估计. 然而, 非参数变量 T_i 不能准确进行测量, 观测到的是具有测量误差扰动的变量 Z_i. 如果不考虑测量误差, 在 profile 非线性最小二乘目标函数 (3.4.4) 中直接用观测变量 Z_i 代替 T_i, 则得到的参数分量和非参数分量的估计不具有相合性. 为了解决这个问题, 需要重新估计 $g(\cdot)$. 借助于 Fan 和 Truong (1993) 提出的思想, 对给定的 β, 利用逆卷积的方法可以得到 $g(\cdot)$ 的估计如下:

$$g_n(t) =: g_n(t;\beta) = \sum_{i=1}^{n} W_{ni}(t)(Y_i - f(X_i,\beta)), \tag{3.4.5}$$

其中

$$W_{ni}(\cdot) =: \frac{1}{nh}K_n\left(\frac{\cdot - Z_j}{h}\right)\Big/\hat{r}_n(\cdot),$$

$$\hat{r}_n(\cdot) = \frac{1}{nh}\sum_{j=1}^{n} K_n\left(\frac{\cdot - Z_j}{h}\right),$$

$$K_n(t) = \frac{1}{2\pi}\int_{-\infty}^{\infty}\exp(-\mathrm{i}st)\frac{\varphi_K(s)}{\phi_e(s/h)}\mathrm{d}s,$$

$\varphi_K(\cdot)$ 是核函数 $K(\cdot)$ 的 Fourier 变换, $\phi_e(\cdot)$ 是误差变量 e 的特征函数. 由式 (3.4.5), 定义如下纠偏的 profile 非线性最小二乘目标函数:

$$Q(\beta) = \sum_{i=1}^{n}\{Y_i - g_n(Z_i;\beta) - f(X_i,\beta)\}^2. \tag{3.4.6}$$

极小化 $Q(\beta)$ 可得 β 的 profile 非线性最小二乘估计 $\hat{\beta}$. 在式 (3.4.5) 中用 $\hat{\beta}$ 代替 β, 则可得到非参数分量 $g(\cdot)$ 的估计为

$$\hat{g}_n(t;\hat{\beta}) = \sum_{i=1}^{n} W_{ni}(t)(Y_i - f(X_i,\hat{\beta})). \tag{3.4.7}$$

进一步, 令 $f_1(t,\beta) = E[f'(X,\beta)|T=t]$, 其中 $f'(X,\beta) = \partial f(X,\beta)/\partial\beta$, 则 $f_1(t,\beta)$ 的估计可定义为

$$\hat{f}_1(t,\hat{\beta}) = \sum_{i=1}^{n} W_{ni}(t) f'(X_i,\hat{\beta}). \tag{3.4.8}$$

3.4.2 条件和渐近性质

首先给出 Fan 和 Truong (1993) 中定义的普通光滑的定义和一些符号概念. 普通光滑分布的定义为以下形式.

α 阶普通光滑 如果误差分布的特征函数 $\phi_e(\cdot)$ 满足

$$d_0|t|^{-\alpha} \leqslant |\phi_e(t)| \leqslant d_1|t|^{-\alpha}, \quad \text{当 } t \to \infty \text{时}, \tag{3.4.9}$$

其中 d_0, d_1, α 为正常数. 例如, Γ 分布 $\dfrac{\gamma^p}{\Gamma(p)}x^{p-1}\exp(-\gamma x)$ 为 $\alpha = p$ 阶普通光滑分布; 双指数分布 $\dfrac{1}{2}\exp(-|x|)$ 为 $\alpha = 2$ 阶普通光滑分布.

令

$$f'(X,\beta) = [\partial f(X,\beta)/\partial\beta]_{p\times 1}, \quad f''(X,\beta) = [\partial^2 f(X,\beta)/\partial\beta\partial\beta^{\mathrm{T}}]_{p\times p}$$

和 $\mathrm{Vech}\{f''(X,\beta)\}$ 是 $f(X,\beta)$ 关于 β 所有二阶导数构成的 $p\times(p+1)/2$ 维向量. 记

$$\frac{\partial f(X_i,\beta)}{\partial\beta_l} = h_l(T_i,\beta) + u_{il}(\beta), \quad 1 \leqslant i \leqslant n, \quad 1 \leqslant l \leqslant p, \tag{3.4.10}$$

其中 $h_l(T_i,\beta) = E\Big(\dfrac{\partial f(X_i,\beta)}{\partial\beta_l}\Big|T_i\Big)$, 则当 $0 < \sup\limits_i E[u_{il}^2(\beta)] < \infty$ 时, 利用大数定律有

$$\begin{aligned} &\frac{1}{n}\sum_{i=1}^n u_i(\beta)u_i^{\mathrm{T}}(\beta) \to \Lambda(\beta), \quad \text{a.s.} \\ &n^{-3/4}\left|\sum_{i=1}^n u_{il}(\beta)\right| = O(1), \quad \text{a.s.} \end{aligned} \tag{3.4.11}$$

对所有 $l = 1,2,\cdots,p$ 均成立, $u_i(\beta) = (u_{i1}(\beta),\cdots,u_{ip}(\beta))^{\mathrm{T}}$, $\Lambda(\beta) = E[f'(X,\beta) - E\{f'(X,\beta)|T\}]^{\otimes 2}$ 是 $p\times p$ 正定矩阵.

对任何向量 a, 用 $\|a\|$ 表示 Euclidean 模. 为了得到本节的主要结论, 需要以下条件.

(C3.4.1) 对 $t \in [0,1]$, $g(t)$, $h_l(t,\beta)$ 和 $E\left(\dfrac{\partial^2 f(X_i,\beta)}{\partial\beta_l\partial\beta_s}\Big|T=t\right)$ 满足 1 阶 Lipschitz 条件, $1 \leqslant l,s \leqslant p$.

(C3.4.2) $0 < \inf\limits_{0\leqslant t\leqslant 1} r(t) \leqslant \sup\limits_{0\leqslant t\leqslant 1} r(t) < \infty$.

(C3.4.3) $E\|u_i\|^4 < \infty$, $\sup\limits_{1\leqslant i\leqslant n} E(\varepsilon_1^4 + \|e_i\|^4) < \infty$.

(C3.4.4) ① 特征函数 $\phi_e(\cdot)$ 非退化; ② 测量误差 e 的分布满足普通光滑条件.

(C3.4.5) 核函数 $K(\cdot)$ 为 k 阶核, 满足:

① $\int_{-\infty}^{\infty} K(u)\mathrm{d}u = 1,\ \int_{-\infty}^{\infty} u^j K(u)\mathrm{d}u = 0, j = 1, \cdots, k-1,\ \int_{-\infty}^{\infty} u^k K(u)\mathrm{d}u \neq 0.$

② $\int_{-\infty}^{\infty} |\varphi_K(t)t^\alpha|\mathrm{d}t < \infty,\ \int_{-\infty}^{\infty} |\varphi_K(t)t^\alpha|^2\mathrm{d}t < \infty$, 其中 α 是 e 的分布的光滑阶数.

(C3.4.6) 对于任意的 x 和 $\beta \in \Theta$, $f(x,\beta)$ 是关于 β 的连续函数, 且 $f(x,\beta)$ 关于 β 的二阶导数是连续的, 其中 Θ 是一个紧集. $E\{f'(X,\beta)\}^{\otimes 2}$ 和 $E[E\{f'(X,\beta)|T\}^{\otimes 2}]$ 在 β_0 的邻域内都有界.

(C3.4.7) $E\{\|f'(X,\beta)\|^4\} < \infty$, $E\{\|\mathrm{Vech}\{f''(X,\beta)\}\|^4\} < \infty$. 进一步, 当 β 在参数真值 β_0 的邻域内时, $\|\mathrm{Vech}\{f''(X,\beta)\}\| \leqslant B(X)$ 和 $E\{\|B(X)\|^4\} < \infty$.

定理 3.4.1 假设条件(C3.4.1)~条件(C3.4.7)成立, 如果取 $h = \rho n^{-1/(2k+2\alpha+1)}$, 其中 $\rho > 0$ 和 $2k > 2\alpha + 1$. 对一些常量 $c \neq 0$, 进一步假设

$$t^\alpha \phi_e(t) \to c,\quad t^{\alpha+1}\phi_e'(t) = O(1),\quad 当\ t \to \infty$$

和

$$\int_{-\infty}^{+\infty} |t|^{\alpha+1}\{|\varphi_K(t)| + |\varphi_K'(t)|\}\mathrm{d}t < \infty,\quad \int_{-\infty}^{+\infty} |t^{\alpha+1}\varphi_K(t)|^2\mathrm{d}t < \infty.$$

则当 $n \to \infty$ 时, $\hat{\beta}$ 是渐近正态的, 即

$$\sqrt{n}(\hat{\beta} - \beta) \xrightarrow{L} N(0, \sigma^2\Lambda^{-1}(\beta)).$$

定理 3.4.1 的证明见 Feng 等 (2014). 由定理 3.4.1, 可以看出 β 的 profile 非线性最小二乘估计 $\hat{\beta}$ 是渐近有效的. 当 T 被准确测量时, $\hat{\beta}$ 的渐近正态结果与 Li 和 Nie (2008) 中的结论相同. 进一步, 为了利用定理 3.4.1 对参数分量 β 进行统计推断, 需要估计渐近方差. 由式 (3.4.7) 和式 (3.4.8), 以及 profile 非线性最小二乘估计 $\hat{\beta}$, 渐近方差 $\sigma^2\Lambda^{-1}(\beta)/n$ 的估计为

$$\frac{\hat{\sigma}^2}{n}\left[\frac{1}{n}\sum_{i=1}^{n}\{f'(X_i,\hat{\beta}) - \hat{f}_1(Z_i,\hat{\beta})\}^{\otimes 2}\right]^{-1}, \tag{3.4.12}$$

其中 $\hat{\sigma}^2 = \frac{1}{n}\sum_{i=1}^{n}\hat{\varepsilon}_i^2$ 和 $\hat{\varepsilon}_i = Y_i - \hat{g}_n(Z_i,\hat{\beta}) - f(X_i,\hat{\beta})$.

定理 3.4.2 假设定理 3.4.1 的条件成立, 当 $n \to \infty$ 时, 则有

$$\sup_{0 \leqslant t \leqslant 1}\left|\hat{g}_n(t;\hat{\beta}) - g(t)\right| = O(\tau_n),\quad \text{a.s.}$$

其中 $\tau_n = \min\{O(n^{-1/3}\log n),\ O(n^{-k/(2k+2\alpha+1)})\}$.

3.4.3 模拟研究

本节通过蒙特卡罗模拟研究在有限样本下对 profile 非线性最小二乘估计方法进行评价. 模拟数据从下面模型产生:

$$\begin{cases} Y_i = \beta_1 X_{i1}/(1+\beta_2 X_{i2}) + g(T_i) + \varepsilon_i, \\ Z_i = T_i + e_i, \end{cases} \quad i = 1, \cdots, n,$$

其中协变量 T_i 服从区间 [0,1] 上的均匀分布, 非参数函数 $g(t) = \sin(\pi t)$, $\beta = (\beta_1, \beta_2)^{\mathrm{T}} = (1.2, -0.75)^{\mathrm{T}}$, $X_{i1} \sim N(0,1)$, $X_{i2} \sim U(0,1)$, 模型误差 $\varepsilon_i \sim N(0, 0.2^2)$, e_i 来自双指数分布:

$$f_e(u) = (\sqrt{2}\sigma_0)^{-1} \exp(-\sqrt{2}|u|/\sigma_0).$$

核函数 $K(\cdot)$ 取如下 4 阶核函数

$$K(x) = \frac{1}{2}(3 - x^2)(\sqrt{2\pi})^{-1} \exp(-x^2/2), 其中|x| < \sqrt{3}.$$

则

$$K_n(x) = (\sqrt{2\pi})^{-1} \exp(-x^2/2) \left\{ \frac{3}{2} - \frac{x^2}{2} - \frac{\sigma_0^2}{2h^2} \left(4x^2 - \frac{5}{2} - \frac{x^4}{2} \right) \right\}.$$

在下面的模拟中, 取样本量大小 n=200, 400 和 600, σ_0^2 取两种情况, 即 1/9 和 1/4. 对每种情况, 重复 1000 次模拟试验. 采用交错核实方法选取窗宽. 为了展示所提方法的完成情况, 比较本节所提出的 profile 非线性最小二乘方法和逆卷积技巧所得到的估计 (PNLS-DT), 与忽略测量误差 e_i 直接使用观测数据 $\{X_i = (X_{i1}, X_{i2})^{\mathrm{T}}, Z_i, Y_i, 1 \leqslant i \leqslant n\}$ 和标准的非线性最小二乘方法得到的估计 (NE). 模拟结果见表 3.4.1~表 3.4.3, 其中表 3.4.1 提供了基于 1000 次重复模拟 β 估计的平均偏差和标准差, 表 3.4.2 报告了基于式 (3.4.12) 所计算的平均渐近方差, 表 3.4.3 报告了 β 置信区间的覆盖概率和平均区间长度, 置信水平为 0.95.

表 3.4.1 β_1 和 β_2 估计的偏差 (Bias) 和标准差 (SD)

方法	n	$\hat{\beta}_1$				$\hat{\beta}_2$			
		$\sigma_0^2 = 1/4$		$\sigma_0^2 = 1/9$		$\sigma_0^2 = 1/4$		$\sigma_0^2 = 1/9$	
		Bias	SD	Bias	SD	Bias	SD	Bias	SD
PNLS-DT	200	0.0083	0.0869	0.0037	0.0671	0.0072	0.0854	−0.0035	0.0648
	400	−0.0052	0.0803	0.0020	0.0460	0.0051	0.0733	−0.0021	0.0516
	600	0.0031	0.0575	−0.0009	0.0322	−0.0027	0.0501	−0.0010	0.0404
NE	200	0.1125	0.1246	0.1002	0.1054	−0.0992	0.0997	0.0781	0.0982
	400	0.0940	0.1057	0.0838	0.0907	0.0784	0.0903	0.0568	0.0865
	600	−0.0761	0.0811	−0.0619	0.0823	−0.0581	0.0782	−0.0497	0.0793

表 3.4.2　估计量的平均渐近方差

方法	n	$\hat{\beta}_1$		$\hat{\beta}_2$	
		$\sigma_0^2=1/4$	$\sigma_0^2=1/9$	$\sigma_0^2=1/4$	$\sigma_0^2=1/9$
PNLS-DT	200	0.0036	0.0023	0.0027	0.0018
	400	0.0021	0.0011	0.0015	0.0010
	600	0.0011	0.0005	0.0009	0.0004
NE	200	0.0116	0.0089	0.0071	0.0042
	400	0.0059	0.0038	0.0040	0.0029
	600	0.0032	0.0022	0.0022	0.0014

表 3.4.3　β_1 和 β_2 置信区间的覆盖概率 (CP) 和平均区间长度 (AL)

方法	n	β_1				β_2			
		$\sigma_0^2=1/4$		$\sigma_0^2=1/9$		$\sigma_0^2=1/4$		$\sigma_0^2=1/9$	
		CP	AL	CP	AL	CP	AL	CP	AL
PNLS-DT	200	0.9060	0.1246	0.9120	0.1087	0.9090	0.1212	0.9140	0.1048
	400	0.9210	0.1085	0.9290	0.0813	0.9230	0.0944	0.9310	0.0796
	600	0.9380	0.0661	0.9420	0.0469	0.9370	0.0685	0.9410	0.0403
NE	200	0.8610	0.2215	0.8770	0.1907	0.8540	0.2136	0.8670	0.1880
	400	0.8870	0.1912	0.8960	0.1544	0.8710	0.1762	0.8920	0.1502
	600	0.9020	0.1243	0.9080	0.1091	0.8890	0.1208	0.9040	0.1016

由表 3.4.1~表 3.4.3, 可以得到以下结论.

(1) 当忽略测量误差时, 得到的参数的粗糙估计 (NE) 是有偏的, 并且偏差随着样本量的变大下降缓慢, 同时偏差随着测量误差的方差增大而增大.

(2) 对于相同的观测样本, 表 3.4.1 给出了三种情况的比较: 测量误差的方差为 $\sigma_0^2=1/9$, $\sigma_0^2=1/4$ 和忽略测量误差时的情形. 可以看出基于本节所提出的方法 (PNLS-DT) 得到的参数分量的估计是渐近无偏的, 与忽略测量误差时得到的估计相比较, 它们有较小的偏差和标准差. 当测量误差的方差为 $\sigma_0^2=1/9$ 时, 所提方法 (PNLS-DT) 完成的最好, 而忽略测量误差时得到的估计最差.

(3) 由表 3.4.2, 可以看出估计量的平均渐近方差随着测量误差方差的增大而变大, 并且可发现本节所提方法是最有效的, 这与所得到的理论结果相一致.

(4) 由表 3.4.3 可以看出, 对于两种不同的测量误差, 当样本量 n 变大时, 两种情况的覆盖概率都趋近于置信水平 0.95, 而区间的平均长度都变小. 对于相同的观测样本, 当测量误差的方差增大时, 覆盖概率变小并远离置信水平 0.95, 而区间的平均长度变大. 同时也发现所提出的方法 (PNLS-DT) 优于忽略测量误差的情形 (NE).

非参数分量的估计 $\hat{g}_n(\cdot)$ 的完成情况将使用如下的均方误差的平方根 (RASE)

来评价

$$\text{RASE} = \left\{ \frac{1}{N} \sum_{s=1}^{N} \left[\hat{g}_n(t_s) - g(t_s) \right]^2 \right\}^{1/2},$$

其中 $t_s(s=1,\cdots,N)$ 为估计函数 $\hat{g}_n(t)$ 的格子点. 在模拟研究中, 取 $N=300$, 并考虑两种不同水平的测量误差方差 $\sigma_0^2=1/9$ 和 $1/4$. 模拟中将本节的结果和另外两种估计进行比较, 即忽略了测量误差 e_i 时得到的非参数分量的粗糙估计 (NE), 和使用精确的测量数据 (即变量 T_i 能够被准确观测) 得到的基准估计 (BE). 基准估计 (BE) 作为一个标准估计或参考估计, 主要来说明所提方法的有效性问题, 即说明所提方法对测量误差的处理的有效性. 表 3.4.4 报告了数值模拟结果, 此外图 3.4.1 展示了样本量取 $n=400$ 时非参数估计 $\hat{g}_n(\cdot)$ 的估计曲线. 由表 3.4.4 和图 3.4.1 可以看出所提方法表现良好.

表 3.4.4 非参数分量估计的 RASE

n	BE	NE	$\sigma_0^2=1/4$	$\sigma_0^2=1/9$
200	0.0919	0.2455	0.1383	0.1164
400	0.0538	0.2012	0.0951	0.0792
600	0.0367	0.1736	0.0728	0.0465

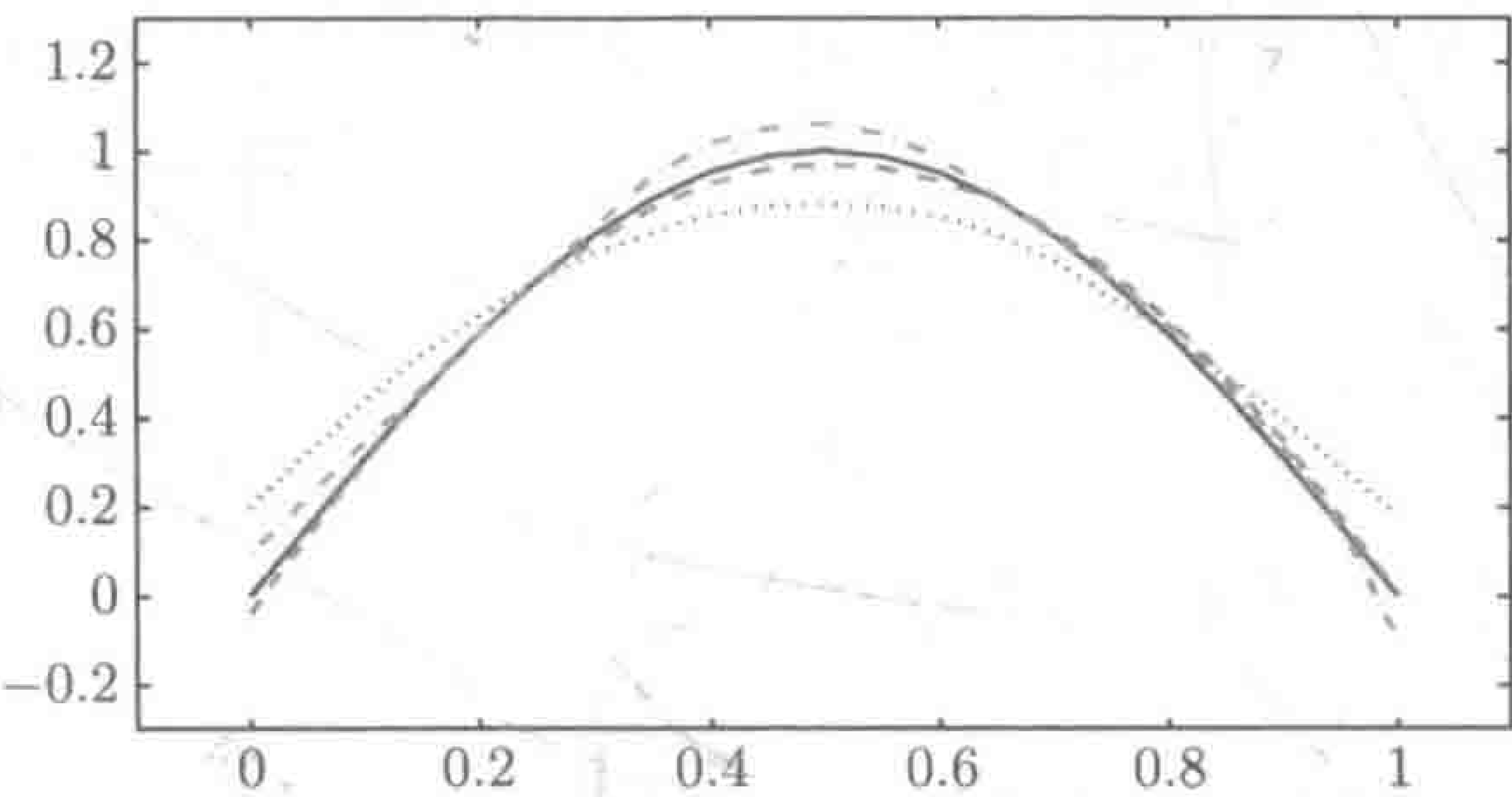


图 3.4.1 非参数函数 $g(\cdot)$ 的估计曲线. 虚线表示 $\sigma_0^2=1/9$ 时的估计曲线, 点断线表示 $\sigma_0^2=1/4$ 时的估计曲线, 点线表示忽略测量误差 e_i 时 (实际方差为 $\sigma_0^2=1/9$) 的估计曲线, 实线表示真实曲线 $g(t)=\sin(\pi t)$

3.5 非线性半参数 EV 模型的经验似然推断

本节利用经验似然方法构造非线性半参数 EV 模型 (3.1.2) 和 (3.1.4) 中未知参数的经验对数似然比统计量, 在适当条件下证明所提出的统计量具有渐近 χ^2 分布, 并利用所得结果构造参数 β 的置信域. 进一步又得到未知参数的极大经验似然

估计, 并证明其渐近正态性. 最后通过模拟研究说明经验似然方法在参数的置信域精度及其覆盖概率大小方面优于最小二乘方法.

3.5.1 方法与主要结果

由式 (3.4.3), 记

$$\widetilde{Y}_i = Y_i - \sum_{j=1}^{n} W_j(T_i)Y_j, \quad \widetilde{f}(X_i,\beta) = f(X_i,\beta) - \sum_{j=1}^{n} W_j(T_i)f(X_j,\beta). \tag{3.5.1}$$

为构造参数分量的经验对数似然比统计量, 首先引入如下的辅助随机变量

$$\eta_i(\beta) = \frac{\partial \widetilde{f}(X_i,\beta)}{\partial\beta}\left[Y_i - f(X_i,\beta) - \hat{g}(T_i)\right] = \frac{\partial \widetilde{f}(X_i,\beta)}{\partial\beta}\left[\widetilde{Y}_i - \widetilde{f}(X_i,\beta)\right]. \tag{3.5.2}$$

当 β 是真实参数时, $E(\eta_i(\beta)) = 0$. 由此可以构造经验对数似然比函数

$$\mathcal{R}(\beta) = -2\max\left\{\sum_{i=1}^{n}\log(np_i) \;\middle|\; p_i \geqslant 0,\ \sum_{i=1}^{n} p_i = 1,\ \sum_{i=1}^{n} p_i\eta_i(\beta) = 0\right\}, \tag{3.5.3}$$

可以证明, $\mathcal{R}(\beta)$ 依分布收敛于一个自由度为 p 的标准 χ^2 分布. 然而, $\mathcal{R}(\beta)$ 还不能直接用来构造 β 的置信域, 因为 T_i 不可观测, 观测到的是 Z_i. 为了解决这个问题, 分别用式 (3.4.2) 中的 $W_{nj}(Z_j)$, $g_n(Z_i)$ 替代式 (3.5.1) 和式 (3.5.2) 中的 $W_j(T_j)$, $\hat{g}(T_i)$, 从而可得估计的辅助随机变量

$$\hat{\eta}_i(\beta) = \frac{\partial \widetilde{f}(X_i,\beta)}{\partial\beta}\left[Y_i - f(X_i,\beta) - g_n(Z_i)\right] = \frac{\partial \widetilde{f}(X_i,\beta)}{\partial\beta}\left[\widetilde{Y}_i - \widetilde{f}(X_i,\beta)\right]. \tag{3.5.4}$$

由此可以构造一个估计的经验对数似然比函数

$$\hat{\mathcal{R}}(\beta) = -2\max\left\{\sum_{i=1}^{n}\log(np_i) \;\middle|\; p_i \geqslant 0,\ \sum_{i=1}^{n} p_i = 1,\ \sum_{i=1}^{n} p_i\hat{\eta}_i(\beta) = 0\right\}. \tag{3.5.5}$$

利用 Lagrange 乘子法, $\hat{\mathcal{R}}(\beta)$ 可以表示为

$$\hat{\mathcal{R}}(\beta) = 2\sum_{i=1}^{n}\log(1 + \lambda^{\mathrm{T}}\hat{\eta}_i(\beta)), \tag{3.5.6}$$

其中 λ 为 Lagrange 乘子, 满足

$$\frac{1}{n}\sum_{i=1}^{n}\frac{\hat{\eta}_i(\beta)}{1 + \lambda^{\mathrm{T}}\hat{\eta}_i(\beta)} = 0. \tag{3.5.7}$$

极大化 $\hat{\mathcal{R}}(\beta)$ 可得 β 的极大经验似然估计 (MELE)$\widetilde{\beta}$. 由 MELE $\widetilde{\beta}$ 及式 (3.4.5) 可以定义 $g(\cdot)$ 的估计为

$$\hat{g}(t)=\sum_{i=1}^{n}W_{ni}(t)(Y_i-f(X_i,\widetilde{\beta})). \tag{3.5.8}$$

定义 σ^2 的估计为

$$\hat{\sigma}^2=\frac{1}{n}\sum_{i=1}^{n}(\tilde{Y}_i-\tilde{f}(X_i,\tilde{\beta}))^2. \tag{3.5.9}$$

定理 3.5.1 假设条件(C3.4.1)~条件(C3.4.6)成立, 取窗宽 $h=d_2n^{-1/3(1+\alpha)}$(d_2 为一正常数). 如果 β 是参数真值, 当 $n\to\infty$ 时, 则有

$$\hat{\mathcal{R}}(\beta)\xrightarrow{L}\chi_p^2.$$

基于定理 3.5.1, 可以定义参数 β 的置信域. 即对任给的 $0<\alpha<1$, 存在 c_α 使得 $P(\chi_p^2>c_\alpha)=\alpha$, 则

$$I_\alpha(\beta^*)=\left\{\beta^*\in\mathbb{R}^p\ \middle|\ \hat{\mathcal{R}}(\beta^*)\leqslant c_\alpha\right\}$$

为参数 β 的具有渐近置信水平 $1-\alpha$ 的置信域.

定理 3.5.2 假设条件(C3.4.1)~条件(C3.4.6)成立, 取窗宽 $h=d_2n^{-1/3(1+\alpha)}$(d_2 为一正常数). 如果 β 是真参数, 当 $n\to\infty$ 时, 则有

$$\sqrt{n}(\widetilde{\beta}-\beta)\xrightarrow{L}N(0,\sigma^2\Lambda(\beta)^{-1}),$$

其中 $\Lambda(\beta)$ 由式 (3.4.11) 定义.

定理 3.5.3 假设条件(C3.4.1)~条件(C3.4.6)成立, 当 $n\to\infty$ 时, 则有

$$\hat{g}(t)-g(t)=O_P(n^{-1/3}\log n).$$

定理 3.5.4 假设条件(C3.4.1)~条件(C3.4.6)成立, 当 $n\to\infty$ 时, 则有

$$\sqrt{n}(\hat{\sigma}^2-\sigma^2)\xrightarrow{L}N(0,\Sigma),$$

其中 $0<\Sigma=:E\varepsilon_1^4-(E\varepsilon_1^2)^2<\infty$.

为了能够将经验似然方法与最小二乘方法进行比较, 给出参数 β 的最小二乘估计 $\hat{\beta}$ 的渐近正态性, 其中 $\hat{\beta}$ 定义为

$$\hat{\beta}=\arg\min_{\beta}\left\{\frac{1}{n}\sum_{i=1}^{n}(Y_i-f(X_i,\beta)-g_n(Z_i))^2\right\}.$$

定理 3.5.5　假设条件(C3.4.1)~条件(C3.4.6)成立, 当 $n \to \infty$ 时, 则有

$$\sqrt{n}(\hat{\beta} - \beta) \xrightarrow{L} N(0, \sigma^2 \Lambda^{-1}(\beta)).$$

由此渐近正态结果也可以用来构造参数 β 的置信域.

限于篇幅, 此处省去了定理 3.5.1~定理 3.5.5 的详细证明. 感兴趣的读者, 可参考冯三营等 (2010a), 冯三营和薛留根 (2012).

3.5.2 模拟研究

本节通过数值模拟计算未知参数 β 的置信域及其覆盖概率的大小, 并比较经验似然 (EL) 方法和最小二乘 (LS) 方法的优劣.

考虑模型:

$$Y = \beta_1(1 - \exp(\beta_2 X)) + \sin(T) + \varepsilon, \quad Z = T + e,$$

其中 $\beta = (1.2, -0.75)^{\mathrm{T}}$, X 服从标准正态分布, T 服从区间 [0,1] 上均匀分布, $\varepsilon \sim N(0, 0.2^2)$. 在下面的模拟中, 取样本容量分别为 n=100, 200, 400, 600. 测量误差 e 服从双指数分布:

$$f_e(x) = (\sqrt{2}/3)^{-1} \exp(-3\sqrt{2}|x|).$$

取核函数为高斯核 $K(u) = (\sqrt{2\pi})^{-1} \exp(-u^2/2)$, 从而

$$K_n(u) = (\sqrt{2\pi})^{-1} \exp(-u^2/2) \left\{1 - \frac{1}{9h^2}(u^2 - 1)\right\}.$$

重复模拟 2000 次. 模拟结果见表 3.5.1 和图 3.5.1.

表 3.5.1　(β_1, β_2) 的置信度为 95% 的置信区间的覆盖概率

n	EL	LS	n	EL	LS
100	0.9060	0.8930	400	0.9430	0.9350
200	0.9290	0.9170	600	0.9480	0.9440

从表 3.5.1 和图 3.5.1, 对比两种方法所得到的模拟结果可以看出, 经验似然 (EL) 方法明显要优于最小二乘方法 (LS), 因为它的覆盖概率大而置信区域小. 另外, 从表 3.5.1 可以看出, 随着样本容量 n 的增大, 两种方法得到的覆盖概率都趋近于置信水平 0.95.

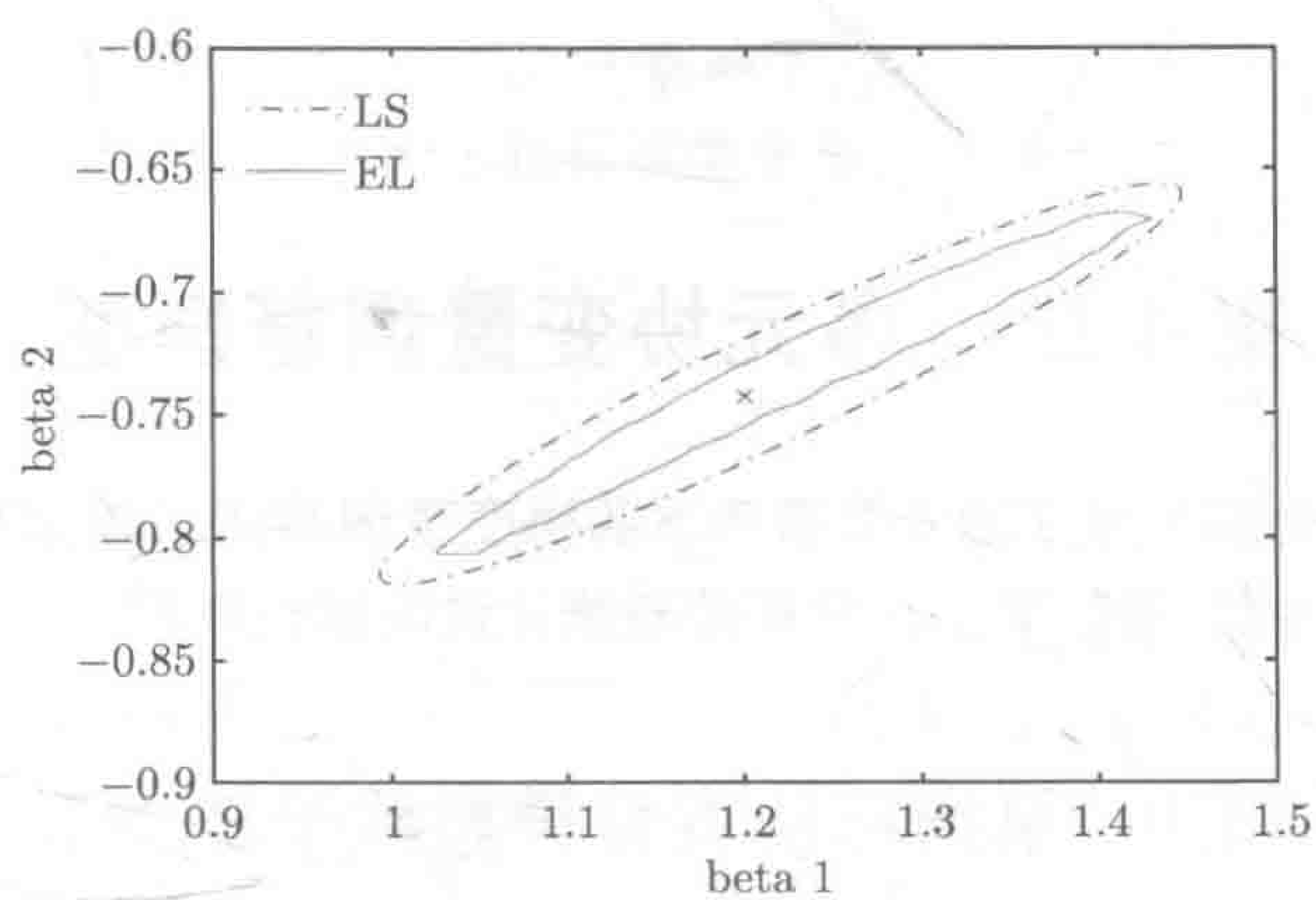


图 3.5.1 $n = 200$ 时, (β_1, β_2) 的置信水平为 0.95 的置信域

第 4 章　多元协变量调整模型

本章将主要介绍多元混淆变量情形下的协变量调整线性模型和非线性模型的估计和统计推断问题, 并研究估计量和检验统计量的渐近性质.

4.1　单指标协变量调整线性模型

4.1.1　引言和模型介绍

在现实生活中, 待研究的变量可能会同时受到多个混淆变量的影响. Nguyen 和 Sentürk (2008) 在研究一组 II-型糖尿病患者数据时发现, 观测到的血糖浓度数据与血管舒张压数据同时被多个混淆变量 (体重指标和三头肌皮褶厚度) 影响. 因此, 需建立多个混淆变量下的统计分析方法. 关于多元协变量调整模型的研究, Nguyen 和 Sentürk (2008) 用单指标模型对失真函数进行建模, 并研究了失真函数中指标参数的估计和线性模型中兴趣参数的估计问题. Zhang 等 (2012a) 考虑了单指标失真函数情形下, 在 Cui 等 (2011) 提出的不动点迭代算法的基础上给出指标参数的估计, 研究了非线性模型中的参数估计问题, 并用经验似然方法构造了非线性模型中兴趣参数的置信域. 不同于 Nguyen 和 Sentürk (2008) 的估计方法, Zhang 等 (2013a) 提出了 "去一分量" 参数估计方法得到了单指标协变量调整线性模型中的指标参数的估计和线性模型中回归系数的估计量, 并研究了估计量的渐近性质. 进一步, Zhang 等 (2014b) 提出将可加模型用于构建失真函数, 研究了非线性模型中参数的估计问题和模型检验问题, 并把经验似然方法用于非线性模型中参数置信域的构造, 理论上研究了估计量和经验对数似然比统计量的渐近性质.

Nguyen 和 Sentürk (2008) 考虑如下的协变量调整线性模型:

$$\begin{cases} Y = \beta_0 + \beta_1 X_1 + \cdots + \beta_p X_p + \varepsilon, \\ \widetilde{Y} = \phi(U) Y, \\ \widetilde{X}_r = \psi_r(U) X_r, \quad r = 1, \cdots, p, \end{cases} \tag{4.1.1}$$

其中 Y 是不可观测的响应变量, $X = (X_1, X_2, \cdots, X_p)^{\mathrm{T}}$ 是不可观测的连续型协变量向量, $\widetilde{Y}$ 和 $\widetilde{X}_r$ 分别是测量到的响应变量和协变量, U 是一个 d 维可观测的混淆向量, $\phi(U)$ 和 $\psi_r(U)$ 是向量 U 的未知的失真函数, $\beta = (\beta_0, \beta_1, \cdots, \beta_p)^{\mathrm{T}}$ 是未知的参数向量.

在多元混淆变量 $U=(U_1,\cdots,U_d)^{\mathrm{T}}$ 影响下, 为了避免 "维数祸根" 问题, Nguyen 和 Sentürk (2008) 首先用如下的单指标模型来拟合失真函数, 即

$$\begin{cases} \phi(U)=\phi(\theta^{\mathrm{T}}U), \\ \psi_r(U)=\psi_r(\theta^{\mathrm{T}}U), \quad r=1,\cdots,p. \end{cases} \tag{4.1.2}$$

在模型 (4.1.2) 中, $\theta=(\theta_1,\cdots,\theta_d)^{\mathrm{T}}$ 是未知的指标参数向量. 不失一般性, 假设 θ 属于参数空间 $\Theta=\{\theta\in\mathbb{R}^d:\|\theta\|=1,\theta_1>0\}$. 失真函数 $\phi(\cdot)$ 和 $\psi_r(\cdot)$ 均是未知的单指标模型的联系函数, $r=1,\cdots,p$. 关于单指标模型更多研究成果的介绍, 读者可参阅 Härdle 等 (1993), Carroll 等 (1997), Zhu 等 (2006), 薛留根 (2012a, 2012b, 2012c) 和 Zhu 等 (2015).

为了介绍估计程序, 需引入如下一些关于模型 (4.1.1) 和模型 (4.1.2) 的基本假设条件:

(1) $E[\phi(\theta^{\mathrm{T}}U)]=1, E[\psi_r(\theta^{\mathrm{T}}U)]=1, r=1,\cdots,p$;

(2) (X,U,ε) 是相互独立的, 且 (Y,U) 也是相互独立的;

(3) $E(\varepsilon)=0, \mathrm{Var}(\varepsilon)=\sigma^2<\infty$.

4.1.2 最小二乘估计方法

假设 $\{(\widetilde{Y}_i,\widetilde{X}_{ir},U_i),i=1,\cdots,n,r=1,\cdots,p\}$ 是来自于模型 (4.1.1) 和模型 (4.1.2) 的 i.i.d.的随机样本. 关于模型 (4.1.2) 中指标参数 θ 的估计方法, Nguyen 和 Sentürk (2008) 注意到不可观测的变量 (Y_i,X_{ir}) 和可观测变量 $(\widetilde{Y}_i,\widetilde{X}_{ir})$ 有如下的关系

$$Y_i=\frac{\widetilde{Y}_i}{\phi(\theta^{\mathrm{T}}U_i)}, \quad X_{ir}=\frac{\widetilde{X}_{ir}}{\psi_r(\theta^{\mathrm{T}}U_i)}, \quad r=1,\cdots,p. \tag{4.1.3}$$

将式 (4.1.3) 代入模型 $Y_i=\beta_0+\beta_1X_{i1}+\cdots+\beta_pX_{ip}+\varepsilon_i$ 中, 可建立如下的单指标变系数模型:

$$\widetilde{Y}_i=\beta_0(\theta^{\mathrm{T}}U_i)+\beta_1(\theta^{\mathrm{T}}U_i)\widetilde{X}_{i1}+\cdots+\beta_p(\theta^{\mathrm{T}}U_i)\widetilde{X}_{ip}+\widetilde{\varepsilon}_i, \tag{4.1.4}$$

其中 $\beta_0(\theta^{\mathrm{T}}U_i)=\phi(\theta^{\mathrm{T}}U_i)\beta_0$, $\beta_r(\theta^{\mathrm{T}}U_i)=\beta_r\phi(\theta^{\mathrm{T}}U_i)/\psi_r(\theta^{\mathrm{T}}U_i)$ 和 $\widetilde{\varepsilon}_i=\phi(\theta^{\mathrm{T}}U_i)\varepsilon_i$, $i=1,\cdots,n,r=1,\cdots,p$.

在模型 (4.1.4) 中变系数函数 $\beta_r(\theta^{\mathrm{T}}u)$ 未知, $r=1,\cdots,p$. 可以采用局部线性光滑估计方法给出系数函数 $\beta_r(\theta^{\mathrm{T}}u)$ 的估计, $r=1,\cdots,p$. 具体算法如下.

步骤 1 首先将 $\beta_r(\theta^{\mathrm{T}}U_i)$ 在 $\theta^{\mathrm{T}}u$ 邻域内进行局部线性 Taylor 展开, 即

$$\begin{aligned}\beta_r(\theta^{\mathrm{T}}U_i)&\approx\beta_r(\theta^{\mathrm{T}}u)+\beta_r'(\theta^{\mathrm{T}}u)(\theta^{\mathrm{T}}U_i-\theta^{\mathrm{T}}u)\\&=:\alpha_r+\alpha_r'(\theta^{\mathrm{T}}U_i-\theta^{\mathrm{T}}u), \quad r=0,1,\cdots,p.\end{aligned} \tag{4.1.5}$$

由式 (4.1.5), 关于 α_r 和 $\alpha'_r(r=0,1,\cdots,p)$ 极小化下面的局部加权最小二乘目标函数

$$\sum_{i=1}^{n}\left[\widetilde{Y}_i-\sum_{r=0}^{p}\left\{\alpha_r+\alpha'_r(\theta^{\mathrm{T}}U_i-\theta^{\mathrm{T}}u)\right\}\widetilde{X}_{ir}\right]^2 K_h(\theta^{\mathrm{T}}U_i-\theta^{\mathrm{T}}u), \tag{4.1.6}$$

其中 $\widetilde{X}_{i0}\equiv 1$, $K_h(\theta^{\mathrm{T}}U_i-\theta^{\mathrm{T}}u)=h^{-1}K\left(\dfrac{\theta^{\mathrm{T}}U_i-\theta^{\mathrm{T}}u}{h}\right)$, $i=1,\cdots,n$, 这里的 h 是窗宽, $K(\cdot)$ 是核密度函数. 把极小化局部加权最小二乘目标函数 (4.1.6) 得到的解记为 $\hat{\alpha}_r=\hat{\beta}_r(\theta^{\mathrm{T}}u), r=0,1,\cdots,p$.

步骤 2　把估计 $\hat{\beta}_r(\theta^{\mathrm{T}}u)$ 代入下面全局的最小二乘目标函数, 则有

$$\mathcal{M}(\theta)=\sum_{i=1}^{n}\left[\widetilde{Y}_i-\sum_{r=0}^{p}\hat{\beta}_r(\theta^{\mathrm{T}}U_i)\widetilde{X}_{ir}\right]^2. \tag{4.1.7}$$

具体地, Nguyen 和 Sentürk (2008) 关于求解目标函数 (4.1.7) 的极小化问题, 得到指标参数 θ 的估计, 提出了如下的迭代算法.

$$\hat{\theta}_{\text{new}}=\hat{\theta}_{\text{old}}-[\mathcal{M}''(\hat{\theta}_{\text{old}})]^{-1}\mathcal{M}'(\hat{\theta}_{\text{old}}),$$

其中$\mathcal{M}''(\hat{\theta}_{\text{old}})=\left.\dfrac{\partial^2\mathcal{M}(\theta)}{\partial\theta\partial\theta^{\mathrm{T}}}\right|_{\theta=\hat{\theta}_{\text{old}}}$, $\mathcal{M}'(\hat{\theta}_{\text{old}})=\left.\dfrac{\partial\mathcal{M}(\theta)}{\partial\theta}\right|_{\theta=\hat{\theta}_{\text{old}}}$, $\hat{\theta}_{\text{old}}$ 是初始估计值或迭代上一步的估计值, $\hat{\theta}_{\text{new}}$ 是更新值. 记最后的收敛值为 $\hat{\theta}^*$, 那么 θ 的估计值为 $\hat{\theta}=\operatorname{sgn}(\hat{\theta}_1^*)\dfrac{\hat{\theta}^*}{\|\hat{\theta}^*\|}$, 其中 $\hat{\theta}_1^*$ 是 $\hat{\theta}^*$ 的第一个分量.

基于模型 (4.1.1), Nguyen 和 Sentürk (2008) 将指标参数 θ 的估计 $\hat{\theta}$ 依次代入极小化目标函数 (4.1.6) 得到的估计 $\hat{\beta}_r(\theta^{\mathrm{T}}u)$ 中, $r=0,1,\cdots,p$. 然后类似于 2.5.3 节中线性模型参数的估计方法, Nguyen 和 Sentürk (2008) 给出了线性模型中兴趣参数如下的矩估计

$$\hat{\beta}_0=\frac{1}{n}\sum_{i=1}^{n}\hat{\beta}_0(\hat{\theta}^{\mathrm{T}}U_i),\qquad \hat{\beta}_r=\frac{1}{\overline{\widetilde{X}}_r}\frac{1}{n}\sum_{i=1}^{n}\hat{\beta}_r(\hat{\theta}^{\mathrm{T}}U_i)\widetilde{X}_{ir},\quad r=1,\cdots,p, \tag{4.1.8}$$

其中 $\overline{\widetilde{X}}_r=\dfrac{1}{n}\sum_{i=1}^{n}\widetilde{X}_{ir}$. Nguyen 和Sentürk (2008) 的数值模拟结果表明估计 (4.1.8) 是模型 (4.1.1) 中系数参数 β_r 的相合估计, $r=0,1,\cdots,p$.

4.1.3 "去一分量" 最小二乘估计方法

1. 估计方法

基于模型 (4.1.1) 和模型 (4.1.2) 的基本假设条件, 以及变量 (Y,X) 和混淆向量

U 的独立性假设, Zhang 等 (2013a) 注意到可根据模型 (4.1.2) 直接建立观测变量 $(\widetilde{Y}, \widetilde{X}_r)$ 与混淆向量 U 之间的单指标模型

$$\begin{cases} E[\widetilde{Y}|U] = \phi(\theta^{\mathrm{T}}U)E(Y) =: g_Y(\theta^{\mathrm{T}}U), \\ E[\widetilde{X}_r|U] = \psi_r(\theta^{\mathrm{T}}U)E(X_r) =: g_{X_r}(\theta^{\mathrm{T}}U), \quad r = 1, \cdots, p. \end{cases} \tag{4.1.9}$$

由模型 (4.1.9), Zhang 等 (2013a) 首先采用局部线性光滑方法给出单指标联系函数 $(g_Y(\theta^{\mathrm{T}}u), g_{X_r}(\theta^{\mathrm{T}}u))$ 和它们导数 $(g'_Y(\theta^{\mathrm{T}}u), g'_{X_r}(\theta^{\mathrm{T}}u))$ 的估计, 即

$$(\hat{g}_Y(\theta^{\mathrm{T}}u), \hat{g}'_Y(\theta^{\mathrm{T}}u)) = \arg\min_{a,b} \sum_{i=1}^n \left[\widetilde{Y}_i - a - b(\theta^{\mathrm{T}}U_i - \theta^{\mathrm{T}}u)\right]^2 K_h(\theta^{\mathrm{T}}U_i - \theta^{\mathrm{T}}u),$$

$$(\hat{g}_{X_r}(\theta^{\mathrm{T}}u), \hat{g}'_{X_r}(\theta^{\mathrm{T}}u)) = \arg\min_{a_r,b_r} \sum_{i=1}^n \left[\widetilde{X}_{ir} - a_r - b_r(\theta^{\mathrm{T}}U_i - \theta^{\mathrm{T}}u)\right]^2 K_h(\theta^{\mathrm{T}}U_i - \theta^{\mathrm{T}}u).$$

进一步, 计算可得

$$\hat{g}_Y(\theta^{\mathrm{T}}u) = \hat{a} = \frac{T_{20}(\theta^{\mathrm{T}}u)T_{01}(\theta^{\mathrm{T}}u) - T_{10}(\theta^{\mathrm{T}}u)T_{11}(\theta^{\mathrm{T}}u)}{T_{20}(\theta^{\mathrm{T}}u)T_{00}(\theta^{\mathrm{T}}u) - T_{10}^2(\theta^{\mathrm{T}}u)},$$

$$\hat{g}_{X_r}(\theta^{\mathrm{T}}u) = \hat{a}_r = \frac{S_{r20}(\theta^{\mathrm{T}}u)S_{r01}(\theta^{\mathrm{T}}u) - S_{r10}(\theta^{\mathrm{T}}u)S_{r11}(\theta^{\mathrm{T}}u)}{S_{r20}(\theta^{\mathrm{T}}u)S_{r00}(\theta^{\mathrm{T}}u) - S_{r10}^2(\theta^{\mathrm{T}}u)},$$

其中

$$T_{l_1 l_2}(\theta^{\mathrm{T}}u) = \sum_{i=1}^n K_h(\theta^{\mathrm{T}}U_i - \theta^{\mathrm{T}}u)(\theta^{\mathrm{T}}U_i - \theta^{\mathrm{T}}u)^{l_1}\widetilde{Y}_i^{l_2},$$

$$S_{rl_1 l_2}(\theta^{\mathrm{T}}u) = \sum_{i=1}^n K_h(\theta^{\mathrm{T}}U_i - \theta^{\mathrm{T}}u)(\theta^{\mathrm{T}}U_i - \theta^{\mathrm{T}}u)^{l_1}\widetilde{X}_{ir}^{l_2},$$

且 $l_1 = 0, 1, 2,\ l_2 = 0, 1,\ r = 1, \cdots, p$.

Zhang 等 (2013a) 定义了如下关于 θ 的目标函数

$$\mathcal{S}(\theta) = \sum_{i=1}^n \left\{ \left[\widetilde{Y}_i - \hat{g}_Y(\theta^{\mathrm{T}}U_i)\right]^2 + \sum_{r=1}^p \left[\widetilde{X}_{ir} - \hat{g}_{X_r}(\theta^{\mathrm{T}}U_i)\right]^2 \right\}.$$

注意到 θ 可重写为 $\theta = \left(\sqrt{1 - \|\theta_{(1)}\|^2}, \theta_{(1)}^{\mathrm{T}}\right)^{\mathrm{T}}$, $\theta_{(1)} = (\theta_2, \cdots, \theta_d)^{\mathrm{T}}$. 求解关于 θ 的估计, 可直接转换为求解关于 $\theta_{(1)}$ 的估计. 为了得到 $\theta_{(1)}$ 的估计, 可以求解下面的估计方程, 即

$$\mathcal{S}'\left(\hat{\theta}_{(1)}\right) = \frac{\partial \mathcal{S}(\theta)}{\partial \theta_{(1)}}\Big|_{\theta_{(1)} = \hat{\theta}_{(1)}} = 0. \tag{4.1.10}$$

对于方程 $\mathcal{S}'\left(\hat{\theta}_{(1)}\right)=0$ 的求解, 可采用类似于 Nguyen 和 Şentürk (2008) 提出的迭代方法, 即有如下的迭代公式

$$\hat{\theta}_{(1)}^{\text{new}}=\hat{\theta}_{(1)}^{\text{old}}-\left[\mathcal{S}''\left(\hat{\theta}_{(1)}^{\text{old}}\right)\right]^{-1}\mathcal{S}'\left(\hat{\theta}_{(1)}^{\text{old}}\right),$$

其中 $\mathcal{S}''\left(\hat{\theta}_{(1)}^{\text{old}}\right)=\left.\dfrac{\partial^2\mathcal{S}(\theta)}{\partial\theta_{(1)}\partial\theta_{(1)}^{\text{T}}}\right|_{\theta_{(1)}=\hat{\theta}_{(1)}^{\text{old}}}$, $\mathcal{S}'\left(\hat{\theta}^{\text{old}}\right)=\left.\dfrac{\partial\mathcal{S}(\theta)}{\partial\theta_{(1)}}\right|_{\theta_{(1)}=\hat{\theta}_{(1)}^{\text{old}}}$, $\hat{\theta}_{(1)}^{\text{old}}$是初始估计值或迭代上一步的估计值, $\hat{\theta}_{(1)}^{\text{new}}$ 是更新值. 记最后的收敛值为 $\hat{\theta}_{(1)}$, 则 θ 的估计值记为 $\hat{\theta}=\left(\sqrt{1-\|\hat{\theta}_{(1)}\|^2},\hat{\theta}_{(1)}^{\text{T}}\right)^{\text{T}}$.

2. 条件和渐近性质

为了得到估计量的渐近结果, 需要下面的假设条件.

(C4.1.1) 失真函数 $\phi(t)$, $\psi_r(t)$, $r=1,\cdots,p$, 随机变量 $\theta^{\text{T}}U$ 的密度函数 $f_{\theta^{\text{T}}U}(t)$ 和函数 $E(U|\theta^{\text{T}}U=t)$ 在集合 $\mathcal{C}$ 上存在三阶连续的导数, 并且三阶导数在 $\mathcal{C}$ 上满足一致 Lipschitz 连续性条件, 其中 $\mathcal{C}=\{t=\theta^{\text{T}}u:u\in\mathcal{U}\subset\mathbb{R}^d\}$, 这里的 $\mathcal{U}$ 是混淆向量 U 的紧支撑.

(C4.1.2) 密度函数 $f_{\theta^{\text{T}}U}(t)$ 在 $t\in\mathcal{C}$ 上恒大于 0. 同时, 失真函数 $\phi(t)$, $\psi_r(t)$, $r=1,\cdots,p$, 在 $t\in\mathcal{C}$ 上不恒为常数且不存在零点.

(C4.1.3) 假设 $E[\phi^2(\theta^{\text{T}}U)]<\infty$ 和 $E[\psi_r^2(\theta^{\text{T}}U)]<\infty$ 成立, $r=1,\cdots,p$.

(C4.1.4) 核函数 $K(t)$ 和 $L(t)$ 是对称有界的一元连续密度函数, 且 $\int t^2K(t)\text{d}t\neq 0$, $\int t^2L(t)\text{d}t\neq 0$, $\int |t|^jK(t)\text{d}t<\infty$, $\int |t|^jL(t)\text{d}t<\infty$, $j=1,2,3$. 此外, $K(t)$ 和 $L(t)$ 的二阶导数有界.

(C4.1.5) $E(X_r)$ 和 $E(Y)$ 都不为 0, $r=1,\cdots,p$.

(C4.1.6) 窗宽 h 满足: 当 $n\to\infty$ 时, $nh^4\to\infty$, $nh^8\to 0$.

(C4.1.7) 窗宽 h_1 满足: 当 $n\to\infty$ 时, $nh_1\to\infty$, $nh_1^4\to 0$.

为表示方便, 定义如下记号, 即

$$\begin{aligned}
\Xi(\theta^{\text{T}}U)=&\left\{\phi'(\theta^{\text{T}}U)E(Y)\right\}^2+\sum_{r=1}^{p}\left\{\psi_r'(\theta^{\text{T}}U)E(X_r)\right\}^2,\\
\Upsilon(\theta^{\text{T}}U)=&\sum_{r=1}^{p}\sum_{t=1}^{p}\left[\phi(\theta^{\text{T}}U)\phi'(\theta^{\text{T}}U)E(Y)\beta_r+\psi_r(\theta^{\text{T}}U)\psi_r'(\theta^{\text{T}}U)E(X_r)\right]\\
&\times\left[\phi(\theta^{\text{T}}U)\phi'(\theta^{\text{T}}U)E(Y)\beta_t+\psi_t(\theta^{\text{T}}U)\psi_t'(\theta^{\text{T}}U)E(X_t)\right]\text{Cov}(X_r,X_t)\\
&+\sigma^2\left\{\phi(\theta^{\text{T}}U)\phi'(\theta^{\text{T}}U)E(Y)\right\}^2.
\end{aligned}$$

定理 4.1.1　假设条件(C4.1.1)~条件(C4.1.6)成立, 当 $n\to\infty$ 时, 则有

$$\sqrt{n}\left(\hat{\theta}_{(1)}-\theta_{(1)}\right) \xrightarrow{L} N(0,\Gamma^{-1}\Lambda\Gamma^{-1}),$$

其中

$$\Gamma = J_{\theta_{(1)}}^{\mathrm{T}} E\big\{\big[E(U|\theta^{\mathrm{T}}U)-U\big]^{\otimes 2}\Xi(\theta^{\mathrm{T}}U)\big\}J_{\theta_{(1)}},$$
$$\Lambda = J_{\theta_{(1)}}^{\mathrm{T}} E\big\{\big[E(U|\theta^{\mathrm{T}}U)-U\big]^{\otimes 2}\Upsilon(\theta^{\mathrm{T}}U)\big\}J_{\theta_{(1)}},$$

且 $J_{\theta_{(1)}} = \partial\theta/\partial\theta_{(1)}$ 是一个 $d\times(d-1)$ 的 Jacobian 矩阵, 定义为

$$J_{\theta_{(1)}} = \begin{pmatrix} -\theta_{(1)}^{\mathrm{T}}\Big/\sqrt{1-\|\theta_{(1)}\|^2} \\ I_{d-1}\end{pmatrix}.$$

由定理 4.1.1 和多元 Delta 定理, 可直接得到如下结果.

定理 4.1.2 在定理 4.1.1 的假设条件下, 当 $n\to\infty$ 时, 则有

$$\sqrt{n}\left(\hat{\theta}-\theta\right) \xrightarrow{L} N(0, J_{\theta_{(1)}}\Gamma^{-1}\Lambda\Gamma^{-1}J_{\theta_{(1)}}^{\mathrm{T}}).$$

下面讨论参数向量 β 的估计. Zhang 等 (2013a) 提出了如下最小二乘方法得到不同于 Nguyen 和 Şentürk (2008) 的估计量, 定义为 $\hat{\beta}_r(\hat{\theta}^{\mathrm{T}}U_i)$, 即有

$$\begin{aligned}&\left(\hat{\beta}_0(\hat{\theta}^{\mathrm{T}}U_i),\hat{\beta}_1(\hat{\theta}^{\mathrm{T}}U_i),\cdots,\hat{\beta}_p(\hat{\theta}^{\mathrm{T}}U_i)\right)\\ &=\arg\min_{b_0\in\mathbb{R},b\in\mathbb{R}^p}\sum_{j=1}^{n}\left[\widetilde{Y}_j-b_0-b^{\mathrm{T}}\widetilde{X}_j\right]^2 L_{h_1}\left(\hat{\theta}^{\mathrm{T}}U_j-\hat{\theta}^{\mathrm{T}}U_i\right),\end{aligned} \tag{4.1.11}$$

其中 $L_{h_1}\left(\hat{\theta}^{\mathrm{T}}U_j-\hat{\theta}^{\mathrm{T}}U_i\right)=h_1^{-1}L\left(\dfrac{\hat{\theta}^{\mathrm{T}}U_j-\hat{\theta}^{\mathrm{T}}U_i}{h_1}\right)$, h_1 是窗宽, $L(\cdot)$ 是核密度函数.

下面的定理给出了由式 (4.1.11) 定义的估计 $\hat{\beta}_r$ 的渐近正态性, $r=0,1,\cdots,p$.

定理 4.1.3 假设条件(C4.1.1)~条件(C4.1.7)成立, 当 $n\to\infty$ 时, 则有

$$\sqrt{n}(\hat{\beta}_r-\beta_r) \xrightarrow{L} N\left(0,\Sigma_{\beta_r}\right),\quad r=0,1,\cdots,p,$$

其中渐近方差 Σ_{β_r} 的表达式为

$$\begin{aligned}\Sigma_{\beta_r} =& e_{r+1,p+1}^{\mathrm{T}}\left\{E\big[(1,X^{\mathrm{T}})^{\mathrm{T}}\big]^{\otimes 2}\right\}^{-1}e_{r,p+1}E\big[\phi^2(\theta^{\mathrm{T}}U)\big]\sigma^2\\ &+\beta_r^2\frac{E(X_r^2)}{(EX_r)^2}E\left\{\big[\phi(\theta^{\mathrm{T}}U)-\psi_r(\theta^{\mathrm{T}}U)\big]^2\right\},\quad r=0,1,\cdots,p,\end{aligned}$$

其中 $\psi_0(\cdot)\equiv 1$, 这里 $e_{r+1,p+1}$ 表示 $p+1$ 维的单位列向量, 第 $r+1$ 个元素为 1, 其他元素全为 0, $r=0,1,\cdots,p$.

定理 4.1.1~定理 4.1.3 的详细证明, 感兴趣的读者请参考 Zhang 等 (2013a).

4.1.4 模拟研究

本节将用数值模拟来说明估计方法的完成情况. 从下面的模型分别产生样本量 $n=200$ 和 $n=400$ 的 500 个数据集.

$$Y=\beta_0+\beta_1X_1+\beta_2X_2+\beta_3X_3+\varepsilon, \tag{4.1.12}$$

其中 $(\beta_0,\beta_1,\beta_2,\beta_3)^{\mathrm{T}}=\left(1,\dfrac{1}{\sqrt{2}},\dfrac{1}{\sqrt{2}},0\right)^{\mathrm{T}}$. 协变量 $(X_1,X_2,X_3)^{\mathrm{T}}$ 来自于多元正态分布 $N_3(\mu_X,0.5^2\Sigma_X)$, 其中 $\mu_X=(2,2,2)^{\mathrm{T}}$, $\Sigma_X=\left(0.5^{|i-j|}\right)_{1\leqslant i,j\leqslant 3}$. 模型误差 ε 来自于正态分布 $N(0,0.25^2)$. 混淆向量 $U=(U_1,U_2,U_3)^{\mathrm{T}}$, 其每一个分量 U_s 都来自区间 $(1,6)$ 的均匀分布, $s=1,2,3$, 且 U 的协方差矩阵为 $\Sigma_U=\left(0.1^{|i-j|}\right)_{1\leqslant i,j\leqslant 3}$, 指标参数 $\theta=\left(\dfrac{1}{\sqrt{3}},\dfrac{1}{\sqrt{3}},\dfrac{1}{\sqrt{3}}\right)^{\mathrm{T}}$. 失真函数选取如下: $\phi(\theta^{\mathrm{T}}U)=\dfrac{(\theta^{\mathrm{T}}U)^2+1}{45.2534}$, $\psi_r(\theta^{\mathrm{T}}U)=\dfrac{\theta^{\mathrm{T}}U+r}{6.4714+r}$, $r=1,2,3$. 核函数取为 Epanechnikov 核函数: $L(t)=K(t)=0.75(1-t^2)I(|t|\leqslant 1)$. 窗宽 h 可由交错核实方法进行选择. 窗宽 h_1 的选取需满足条件 (C4.1.7) 中的欠光滑条件, 这里采用 $h_1=\hat{\sigma}_{\hat{\theta}^{\mathrm{T}}U}n^{-1/3}$, 其中

$$\hat{\sigma}_{\hat{\theta}^{\mathrm{T}}U}=\sqrt{n^{-1}\sum_{i=1}^{n}\{\hat{\theta}^{\mathrm{T}}U_i-\overline{\hat{\theta}^{\mathrm{T}}U}\}^2},\quad \overline{\hat{\theta}^{\mathrm{T}}U}=n^{-1}\sum_{i=1}^{n}\hat{\theta}^{\mathrm{T}}U_i.$$

模拟结果报告在表 4.1.1 和表 4.1.2 中, 其中表 4.1.1 提供了指标参数 θ 估计情况, 表 4.1.2 比较了 Zhang 等 (2013a) 提出的 4.1.3 节介绍的估计方法 (用 “Z” 表示), Sentürk 和 Müller (2006) 提出的 binning 估计方法 (用 “B” 表示) 和 Cui 等 (2009) 提出的直接估计方法 (用 “C” 表示). 另外, 表中 “Mean” 表示估计的均值, “SD” 表示估计的标准差.

表 4.1.1　指标参数 θ 估计的均值 (Mean)、标准差 (SD) 和夹角

n		θ_1	θ_2	θ_3	夹角
$n=200$	Mean	0.5773	0.5761	0.5779	0.0288
	SD	0.0177	0.0180	0.0178	0.0111
$n=400$	Mean	0.5767	0.5761	0.5782	0.0215
	SD	0.0161	0.0166	0.0158	0.0104

表 4.1.2　参数 β 估计的均值 (Mean) 和标准差 (SD)

n	方法		β_0	β_1	β_2	β_3
$n=200$	Z	Mean	0.9639	0.7218	0.7083	0.0079
		SD	0.1348	0.0571	0.0709	0.0492

续表

n	方法		β_0	β_1	β_2	β_3
$n=200$	B	Mean	0.8471	0.7496	0.7205	0.0226
		SD	0.1508	0.0662	0.0750	0.0645
	C	Mean	0.9947	0.6998	0.6991	0.0001
		SD	0.1014	0.0435	0.0490	0.0420
$n=400$	Z	Mean	0.9802	0.7154	0.7119	0.0100
		SD	0.0649	0.0358	0.0391	0.0362
	B	Mean	0.8459	0.7433	0.7233	0.0234
		SD	0.1031	0.0431	0.0488	0.0439
	C	Mean	0.9909	0.6969	0.6994	0.0002
		SD	0.0706	0.0318	0.0335	0.0295

从表 4.1.1 和表 4.1.2 可以看出如下的结果.

(1) 从表 4.1.1 中可以看出, 指标参数 θ 的估计 $\hat{\theta}$ 与真值 θ 的大小都比较接近. 此外, 随着样本量的增加, 估计 $\hat{\theta}$ 与真值 θ 的平均夹角 (以弧度表示) 及其标准差越来越接近于 0;

(2) 从表 4.1.2 中可以看出, Zhang 等 (2013a) 提出的估计和 Cui 等 (2009) 提出的直接估计方法所得估计随着样本量的增加, 估计表现效果较好. 而 binning 估计方法得到的关于截距项 β_0 的估计精度较差, 原因在于 binning 估计方法是比较粗糙的一种非参数光滑方法.

4.2 单指标协变量调整非线性模型

4.2.1 模型介绍

本节考虑下面的单指标协变量调整非线性模型:

$$\begin{cases} Y = f(X,\beta) + \varepsilon, \\ \widetilde{Y} = \phi(\theta^{\mathrm{T}}U)Y, \\ \widetilde{X}_r = \psi_r(\theta^{\mathrm{T}}U)X_r, \quad r = 1,\cdots,p, \end{cases} \tag{4.2.1}$$

其中 Y 是不可观测的响应变量, $f(X,\beta)$ 是函数形式已知的非线性函数, β 是未知的 q 维参数向量, $X=(X_1,X_2,\cdots,X_p)^{\mathrm{T}}$ 是不可观测的连续型协变量, $\widetilde{Y}$ 和 $\widetilde{X}_r(r=1,\cdots,p)$ 是可观测的变量, U 是一个可观测的 d 维混淆向量, $\phi(\theta^{\mathrm{T}}U)$ 和 $\psi_r(\theta^{\mathrm{T}}U)$ 是混淆向量 U 的未知单指标失真函数, 指标参数 $\theta\in\Theta$, 其中 $\Theta=\{\theta=(\theta_1,\theta_2,\cdots,\theta_d)^{\mathrm{T}}: \|\theta\|=1,\theta_1>0,\theta\in\mathbb{R}^d\}$.

为了介绍估计程序, 需要引入如下一些关于模型 (4.2.1) 的基本假设条件:

(1) $E[\phi(\theta^{\mathrm{T}}U)]=1, E[\psi_r(\theta^{\mathrm{T}}U)]=1, r=1,\cdots,p$;

(2) (X,U,ε) 是相互独立的, 且 (Y,U) 也是相互独立的;

(3) $E(\varepsilon)=0, \mathrm{Var}(\varepsilon)=\sigma^2<\infty$.

4.2.2 指标参数 θ 的估计

基于混淆向量 U 和变量 (Y,X) 的独立性假设条件, 对 $r=1,\cdots,p$, 可得到下面的关系式

$$E(\widetilde{Y}|U)=\phi(\theta^{\mathrm{T}}U)E(Y),\quad \mathrm{Var}(\widetilde{Y}|U)=\mathrm{Var}(Y)\phi^2(\theta^{\mathrm{T}}U), \tag{4.2.2}$$

$$E(\widetilde{X}_r|U)=\psi_r(\theta^{\mathrm{T}}U)E(X_r),\quad \mathrm{Var}(\widetilde{X}_r|U)=\mathrm{Var}(X_r)\psi_r^2(\theta^{\mathrm{T}}U). \tag{4.2.3}$$

从式 (4.2.2) 和式 (4.2.3) 可以看出, 条件均值和条件方差均是单指标模型. 这时可直接使用 Cui 等 (2011) 提出的不动点迭代估计方法给出指标参数 θ 的估计. 关于不动点迭代方法, 有兴趣的读者可参阅文献 Cui 等 (2011).

假定有 i.i.d. 的观测样本 $\{(\widetilde{X}_i,\widetilde{Y}_i,U_i), i=1,\cdots,n\}$ 来自于模型 (4.2.1). 先对 $\phi(\theta^{\mathrm{T}}u)$ 和 $\psi_r(\theta^{\mathrm{T}}u), r=1,\cdots,p$, 在 t 处进行局部线性逼近, 即

$$\begin{aligned}\phi(\theta^{\mathrm{T}}u)&\approx\gamma_0+\gamma_1(\theta^{\mathrm{T}}u-t)=:\phi_a(t),\\ \psi_r(\theta^{\mathrm{T}}u)&\approx\gamma_{r0}+\gamma_{r1}(\theta^{\mathrm{T}}u-t)=:\psi_{ar}(t),\quad r=1,\cdots,p.\end{aligned}$$

在给定指标参数 θ 时, 不动点迭代方法首先需要构建如下 $p+1$ 个局部估计方程组, 求解方程组来得到 (γ_0,γ_1) 和 $(\gamma_{r0},\gamma_{r1})$ 的估计. 方程组定义如下:

$$\begin{cases}\displaystyle\sum_{i=1}^n K_h(\theta^{\mathrm{T}}U_i-t)[\phi_a^2(\theta^{\mathrm{T}}U_i)]^{-1}\left(\widetilde{Y}_i-\overline{\widetilde{Y}}\phi_a(\theta^{\mathrm{T}}U_i)\right)=0,\\ \displaystyle\sum_{i=1}^n K_h(\theta^{\mathrm{T}}U_i-t)(\theta^{\mathrm{T}}U_i-t)[\phi_a^2(\theta^{\mathrm{T}}U_i)]^{-1}\left(\widetilde{X}_{ir}-\overline{\widetilde{Y}}\phi_a(\theta^{\mathrm{T}}U_i)\right)=0\end{cases} \tag{4.2.4}$$

和

$$\begin{cases}\displaystyle\sum_{i=1}^n K_h(\theta^{\mathrm{T}}U_i-t)[\psi_{ar}^2(\theta^{\mathrm{T}}U_i)]^{-1}\left(\widetilde{X}_{ir}-\overline{\widetilde{X}}_r\psi_{ar}(\theta^{\mathrm{T}}U_i)\right)=0,\\ \displaystyle\sum_{i=1}^n K_h(\theta^{\mathrm{T}}U_i-t)(\theta^{\mathrm{T}}U_i-t)[\psi_{ar}^2(\theta^{\mathrm{T}}U_i)]^{-1}\left(\widetilde{X}_{ir}-\overline{\widetilde{X}}_r\psi_{ar}(\theta^{\mathrm{T}}U_i)\right)=0,\end{cases} \tag{4.2.5}$$

其中 $\overline{\widetilde{Y}}=\dfrac{1}{n}\displaystyle\sum_{i=1}^n\widetilde{Y}_i$, $\overline{\widetilde{X}}_r=\dfrac{1}{n}\displaystyle\sum_{i=1}^n\widetilde{X}_{ir}$, $K_h(\cdot)=h^{-1}K(\cdot/h)$, $K(\cdot)$ 是核密度函数, h 是窗宽. 对方程 (4.2.4) 和方程 (4.2.5) 关于 $(\gamma_0,\ \gamma_1)$ 和 $(\gamma_{r0},\ \gamma_{r1})$ 进行求解, 可得

到 $\phi(t)$ 和 $\psi_r(t)$ 在 t 点处的估计, 定义为 $\hat{\phi}(t)=\hat{\gamma}_0$, $\hat{\phi}'(t)=\hat{\gamma}_1$, $\hat{\psi}_r(t)=\hat{\gamma}_{r0}$ 和 $\hat{\phi}'_r(t)=\hat{\gamma}_{r1}, r=1,\cdots,p$.

注意到 θ 可重写为 $\theta=\left(\sqrt{1-\|\theta_{(1)}\|^2},\theta_{(1)}^{\mathrm{T}}\right)^{\mathrm{T}}$, 其中 $\theta_{(1)}=(\theta_2,\cdots,\theta_d)^{\mathrm{T}}$. 可以构建如下 $p+1$ 个关于 $\theta_{(1)}$ 的拟似然估计方程: 对 $r=1,\cdots,p$, 有

$$\hat{\Phi}_r(\theta_{(1)})=\sum_{i=1}^n J_{\theta_{(1)}}^{\mathrm{T}}\hat{\psi}'_r(\theta^{\mathrm{T}}U_i)\big(U_i-\hat{s}(\theta^{\mathrm{T}}U_i)\big)\big[\hat{\psi}_r^2(\theta^{\mathrm{T}}U_i)\big]^{-1}\left(\widetilde{X}_{ir}-\overline{\widetilde{X}}_r\hat{\psi}_r(\theta^{\mathrm{T}}U_i)\right)=0 \quad (4.2.6)$$

和

$$\hat{\Phi}_{p+1}(\theta_{(1)})=\sum_{i=1}^n J_{\theta_{(1)}}^{\mathrm{T}}\hat{\phi}'(\theta^{\mathrm{T}}U_i)\big(U_i-\hat{s}(\theta^{\mathrm{T}}U_i)\big)\big[\hat{\phi}^2(\theta^{\mathrm{T}}U_i)\big]^{-1}\left(\widetilde{Y}_i-\overline{\widetilde{Y}}\hat{\phi}(\theta^{\mathrm{T}}U_i)\right)=0, \quad (4.2.7)$$

其中 $J_{\theta_{(1)}}=\partial\theta/\partial\theta_{(1)}$ 是 $d\times(d-1)$ 的 Jacobian 矩阵, 定义为

$$J_{\theta_{(1)}}=\begin{pmatrix}-\theta_{(1)}^{\mathrm{T}}/\sqrt{1-\|\theta_{(1)}\|^2}\\ I_{d-1}\end{pmatrix}.$$

方程 (4.2.6) 和方程 (4.2.7) 中的 $\hat{s}(t)$ 是 $s(t)=E(U|\theta^{\mathrm{T}}U=t)=(s_1(t),\cdots,s_d(t))^{\mathrm{T}}$ 的局部线性光滑估计, 定义为

$$\hat{s}(t)=\frac{\sum_{i=1}^n b_i(t)U_i}{\sum_{i=1}^n b_i(t)},$$

其中

$$b_i(t)=K_h(\theta^{\mathrm{T}}U_i-t)\big[S_{n,2}(t)-(\theta^{\mathrm{T}}U_i-t)S_{n,1}(t)\big],$$
$$S_{n,j}=\sum_{i=1}^n K_h(\theta^{\mathrm{T}}U_i-t)(\theta^{\mathrm{T}}U_i-t)^j, j=1,2.$$

记 $\hat{\theta}_{(1)}^{[r]}$ 为第 r 个估计方程 $\hat{\Phi}_r\left(\hat{\theta}_{(1)}^{[r]}\right)=0$ 得到的参数 $\theta_{(1)}$ 的估计. 这时, $\theta_{(1)}$ 的最终估计定义为

$$\hat{\theta}_{(1)}=\frac{1}{p+1}\sum_{r=1}^{p+1}\hat{\theta}_{(1)}^{[r]}. \quad (4.2.8)$$

得到参数 $\theta_{(1)}$ 的估计 $\hat{\theta}_{(1)}$ 以后, 则可以定义指标参数向量 θ 第一个分量 θ_1 的估计为 $\hat{\theta}_1=\sqrt{1-\|\hat{\theta}_{(1)}\|^2}$. 因此, 指标参数向量 θ 的估计可以定义为 $\hat{\theta}=\left(\hat{\theta}_1,\hat{\theta}_{(1)}^{\mathrm{T}}\right)^{\mathrm{T}}$.

4.2.3 参数 β 的估计

关于参数 β 的估计, 可以通过如下三步估计方法得到, 即: ① 采用直接估计方法先把未知的失真函数估计出来; ② 把不可观测的协变量与响应变量估计出来; ③ 给出参数 β 的估计.

注意到

$$E\left\{\frac{\widetilde{Y}}{E(Y)}\middle|\theta^{\mathrm{T}}U\right\}=\phi(\theta^{\mathrm{T}}U),\quad E\left\{\frac{\widetilde{X}_r}{E(X_r)}\middle|\theta^{\mathrm{T}}U\right\}=\psi_r(\theta^{\mathrm{T}}U),\quad r=1,\cdots,p.$$

进一步, 可用局部线性光滑方法重新给出 $\phi(\cdot)$ 和 $\psi_r(\cdot)$ 的估计, 定义为

$$\hat{\phi}_b(t)=\frac{\sum_{i=1}^{n}r_i(t,\hat{\theta})\widetilde{Y}_i}{\sum_{i=1}^{n}r_i(t,\hat{\theta})\overline{\widetilde{Y}}},\quad \hat{\psi}_{br}(t)=\frac{\sum_{i=1}^{n}r_i(t,\hat{\theta})\widetilde{X}_{ir}}{\sum_{i=1}^{n}r_i(t,\hat{\theta})\overline{\widetilde{X}}_r},\quad r=1,\cdots,p,$$

其中

$$r_i(t,\hat{\theta})=L_{h_1}(\hat{\theta}^{\mathrm{T}}U_i-t)\left[Q_{n,2}(t,\hat{\theta})-(\hat{\theta}^{\mathrm{T}}U_i-t)Q_{n,1}(t,\hat{\theta})\right],$$

$$Q_{n,j}(t,\hat{\theta})=\sum_{i=1}^{n}L_{h_1}(\hat{\theta}^{\mathrm{T}}U_i-t)(\hat{\theta}^{\mathrm{T}}U_i-t)^j,j=1,2.$$

这里 $L_{h_1}(\cdot)=h_1^{-1}L(\cdot/h_1)$, $L(\cdot)$ 是一个核密度函数, h_1 是窗宽.

因此, 通过关系式 $Y=\dfrac{\widetilde{Y}}{\phi(\theta^{\mathrm{T}}U)}$ 和 $X_r=\dfrac{\widetilde{X}_r}{\psi_r(\theta^{\mathrm{T}}U)}$ 来构造恢复数据 $\{(\hat{Y}_i,\hat{X}_i),i=1,\cdots,n\}$, 定义为

$$\hat{Y}_i=\frac{\widetilde{Y}_i}{\hat{\phi}_b(\hat{\theta}^{\mathrm{T}}U_i)},\quad \hat{X}_{ir}=\frac{\widetilde{X}_{ir}}{\hat{\psi}_{br}(\hat{\theta}^{\mathrm{T}}U_i)},\quad r=1,\cdots,p,\quad i=1,\cdots,n. \tag{4.2.9}$$

基于恢复数据 $\{(\hat{Y}_i,\hat{X}_i),i=1,\cdots,n\}$, 可由非线性最小二乘方法来得到 β 的估计 $\hat{\beta}$, 该估计可通过求解下面的 q 个估计方程得到

$$\sum_{i=1}^{n}\left(\hat{Y}_i-f(\hat{X}_i,\beta)\right)\frac{\partial f(\hat{X}_i,\beta)}{\partial\beta_k}=0,\quad k=1,\cdots,q, \tag{4.2.10}$$

其中 $\partial f(\cdot,\beta)/\partial\beta_k$ 是 $f(\cdot,\beta)$ 关于第 k 个分量 β_k 的偏导数. 若估计方程 (4.2.10) 没有显示表达式, 可用 Newton-Raphson 迭代算法来得到估计 $\hat{\beta}$ 的数值解. 把解估计方程 (4.2.10) 得到的估计称为非线性最小二乘估计, 记为 $\hat{\beta}$.

4.2.4 条件和渐近性质

本节给出估计量 $\hat{\theta}$ 和 $\hat{\beta}$ 的渐近性质. 为了给出估计量的渐近性质, 首先列出定理所需要的技术条件.

(C4.2.1) 失真函数 $\phi(t)$, $\psi_r(t)$, $r=1,\cdots,p$, 随机变量 $\theta^{\mathrm{T}}U$ 的密度函数 $f_{\theta^{\mathrm{T}}U}(t)$ 和函数 $E(U|\theta^{\mathrm{T}}U=t)$ 在集合 $\mathcal{C}$ 上存在三阶连续的导数, 且三阶导数在 $\mathcal{C}$ 上满足一致 Lipschitz 连续性条件, 其中 $\mathcal{C}=\{t=\theta^{\mathrm{T}}u:u\in\mathcal{U}\subset\mathbb{R}^d\}$, 这里的 $\mathcal{U}$ 是混淆向量 U 的一个紧支撑.

(C4.2.2) 密度函数 $f_{\theta^{\mathrm{T}}U}(t)$ 在 $t \in \mathcal{C}$ 上恒大于 0. 同时, 失真函数 $\phi(t)$, $\psi_r(t)$, $r=1,\cdots,p$, 在 $t \in \mathcal{C}$ 上不恒为常数且不存在零点.

(C4.2.3) 核函数 $K(t)$ 和 $L(t)$ 是对称有界的一元连续密度函数, 满足 $\int t^2K(t)\mathrm{d}t \neq 0$, $\int t^2L(t)\mathrm{d}t \neq 0$, $\int |t|^jK(t)\mathrm{d}t < \infty$, $\int |t|^jL(t)\mathrm{d}t < \infty$, $j=1,2,3$. 此外, $K(t)$ 和 $L(t)$ 的二阶导数有界.

(C4.2.4) 窗宽 h 满足: 当 $n \to \infty$ 时, $nh^4 \to \infty$, $nh^6 \to 0$.

(C4.2.5) $E(X_r) \neq 0$ 和 $E(Y) \neq 0$, $r=1,\cdots,p$.

(C4.2.6) 模型误差 ε 满足 $E\varepsilon^4 < \infty$.

(C4.2.7) 窗宽 h_1 满足: 当 $n \to \infty$ 时, $\dfrac{(\log n)^2}{nh_1^2} \to 0$, $nh_1^4 \to 0$.

(C4.2.8) 当 $l_1,l_2,l_3,l_4=0,1,2$, $l_1+l_2+l_3+l_4 \leqslant 3$, $1 \leqslant s_1,s_2 \leqslant p$ 和 $1 \leqslant t_1,t_2 \leqslant q$ 时, 偏导数

$$\frac{\partial^{l_1+l_2+l_3+l_4} f(X,\beta)}{\partial^{l_1}\beta_{t_1}\partial^{l_2}\beta_{t_2}\partial^{l_3}X_{s_1}\partial^{l_4}X_{s_2}}$$

存在, 并存在正数 C_1 使得下面的不等式

$$\left|\frac{\partial^{l_1+l_2+l_3+l_4} f(X,\beta)}{\partial^{l_1}\beta_{t_1}\partial^{l_2}\beta_{t_2}\partial^{l_3}X_{s_1}\partial^{l_4}X_{s_2}}\right| \leqslant C_1$$

在 $l_3+l_4 \geqslant 1$ 时成立. 同时存在正数 C_2 使得下面的不等式

$$E\left\{\sup_{\beta}\left|\frac{\partial^{l_1+l_2+l_3+l_4} f(X,\beta)}{\partial^{l_1}\beta_{t_1}\partial^{l_2}\beta_{t_2}\partial^{l_3}X_{s_1}\partial^{l_4}X_{s_2}}\right|\right\} \leqslant C_2$$

在 $1 \leqslant l_1+l_2 \leqslant 2$ 和 $l_3+l_4=0$ 时成立.

(C4.2.9) 真值 β 满足: $\beta = \arg\min\limits_{\beta^* \in \mathbb{R}^q} E\big[f(X,\beta)-f(X,\beta^*)\big]^2$.

定理 4.2.1 假设条件(C4.2.1)~条件(C4.2.7)成立, 当 $n \to \infty$ 时, 则有

$$\sqrt{n}(\hat{\theta}-\theta) \xrightarrow{L} N\left(0, \frac{1}{(p+1)^2}J_{\theta_{(1)}}\Sigma_\theta J_{\theta_{(1)}}^{\mathrm{T}}\right),$$

其中 Σ_θ 定义为

$$\begin{aligned}\Sigma_\theta = {} & E\left\{\sum_{s=1}^{p}\sum_{t=1}^{p}(EX_sEX_t)\Gamma_s^{-1}J_{\theta_{(1)}}^{\mathrm{T}}\breve{U}^{\otimes 2}J_{\theta_{(1)}}\Gamma_t^{-1}\frac{\psi_s'(\theta^{\mathrm{T}}U)\psi_t'(\theta^{\mathrm{T}}U)}{\psi_s(\theta^{\mathrm{T}}U)\psi_t(\theta^{\mathrm{T}}U)}\mathrm{Cov}(X_s,X_t)\right\}\\ & +E\left\{\sum_{s=1}^{p}(EYEX_s)(\Gamma_s^{-1}J_{\theta_{(1)}}^{\mathrm{T}}\breve{U}^{\otimes 2}J_{\theta_{(1)}}\Gamma_{p+1}^{-1}+\Gamma_{p+1}^{-1}J_{\theta_{(1)}}^{\mathrm{T}}\breve{U}^{\otimes 2}J_{\theta_{(1)}}\Gamma_s^{-1})\right.\\ & \left.\times\frac{\psi_s'(\theta^{\mathrm{T}}U)\phi'(\theta^{\mathrm{T}}U)}{\psi_s(\theta^{\mathrm{T}}U)\phi(\theta^{\mathrm{T}}U)}\mathrm{Cov}(X_s,Y)\right\}+\mathrm{Var}(Y)\Gamma_{p+1}^{-1},\end{aligned}$$

这里 $\breve{U}=U-E(U|\theta^{\mathrm{T}}U)$. 此外, $\Gamma_r(r=1,\cdots,p)$ 和 Γ_{p+1} 的定义分别为

$$\Gamma_r=(EX_r)^2E\left(\frac{\psi_r{}'(\theta^{\mathrm{T}}U)}{\psi_r(\theta^{\mathrm{T}}U)}\right)^2J_{\theta_{(1)}}^{\mathrm{T}}\breve{U}^{\otimes 2}J_{\theta_{(1)}},$$

$$\Gamma_{p+1}=(EY)^2E\left(\frac{\phi'(\theta^{\mathrm{T}}U)}{\phi(\theta^{\mathrm{T}}U)}\right)^2J_{\theta_{(1)}}^{\mathrm{T}}\breve{U}^{\otimes 2}J_{\theta_{(1)}}.$$

定理 4.2.2　假设条件(C4.2.1)~条件(C4.2.9)成立, 当 $n\to\infty$ 时, 则有

(1) $\hat{\beta}$ 依概率 1 收敛到真值 β;

(2) β 的估计 $\hat{\beta}$ 具有渐近正态分布, 即

$$\sqrt{n}(\hat{\beta}-\beta)\xrightarrow{L}N(0,\Lambda^{-1}\Omega\Lambda^{-1}),$$

其中矩阵 Λ 与矩阵 Ω 的第 (s,t) 个位置上的元素为 $\Omega(s,t)=\sigma^2\Lambda(s,t)+\Upsilon(s,t)$, 且

$$\begin{aligned}
\Lambda(s,t)=&E\left(\frac{\partial f(X,\beta)}{\partial\beta_s}\frac{\partial f(X,\beta)}{\partial\beta_t}\right),\\
\Upsilon(s,t)=&\mathrm{Var}\left(\frac{\widetilde{Y}-Y}{EY}\right)E\left(Y\frac{\partial f(X,\beta)}{\partial\beta_s}\right)E\left(Y\frac{\partial f(X,\beta)}{\partial\beta_t}\right)\\
&+\sum_{r=1}^{p}\sum_{l=1}^{p}\mathrm{Cov}\left(\frac{\widetilde{X}_r-X_r}{EX_r},\frac{\widetilde{X}_l-X_l}{EX_l}\right)E\left(X_r\frac{\partial f(X,\beta)}{\partial X_r}\frac{\partial f(X,\beta)}{\partial\beta_s}\right)\\
&\times E\left(X_l\frac{\partial f(X,\beta)}{\partial X_l}\frac{\partial f(X,\beta)}{\partial\beta_t}\right)\\
&-\sum_{r=1}^{p}\mathrm{Cov}\left(\frac{\widetilde{X}_r-X_r}{EX_r},\frac{\widetilde{Y}-Y}{EY}\right)\left\{E\left(X_r\frac{\partial f(X,\beta)}{\partial X_r}\frac{\partial f(X,\beta)}{\partial\beta_s}\right)\right.\\
&\times E\left(Y\frac{\partial f(X,\beta)}{\partial\beta_t}\right)\\
&\left.+E\left(X_r\frac{\partial f(X,\beta)}{\partial X_r}\frac{\partial f(X,\beta)}{\partial\beta_t}\right)E\left(Y\frac{\partial f(X,\beta)}{\partial\beta_s}\right)\right\}.
\end{aligned}$$

4.2.5　经验似然置信域

关于参数 β 的置信域, 可根据定理 4.2.2 的渐近正态结果, 把 $\hat{\beta}$ 的渐近方差中的每一个元素估计出来, 然后用渐近正态分布近似的方法给出 β 的置信区间 $I_{\alpha,\mathrm{NOR}}=\{\beta^*:n(\hat{\beta}-\beta^*)^{\mathrm{T}}\hat{\Sigma}^{-1}(\hat{\beta}-\beta^*)\leqslant c_\alpha\}$, 其中 $\hat{\Sigma}=\hat{\Lambda}^{-1}\hat{\Omega}\hat{\Lambda}^{-1}$ 是定理 4.2.2 中渐近方差的一个 plug-in 相合估计. 当样本量比较大的时候, 可直接利用渐近正态分布近似得到置信区间, 因为在大样本情形下的 plug-in 相合估计往往表现较好. 但是在样本量较小的时候, 不是一个很好的选择, 因为此时构造的 plug-in 估计往往表现较差. 为了能够得到更精确的置信域, 可以采用由 Owen (1988, 1990, 1991) 提出的经

验似然方法来构造兴趣参数 β 的置信域. 关于经验似然方法构造置信域的优点和详细讨论, 可参考 1.3 节.

下面详细介绍构造 β 的置信域的经验似然方法.

首先引入一个辅助随机向量 $\varpi_{n,i}(\beta^*)=(\varpi_{n,i}^1(\beta^*),\cdots,\varpi_{n,i}^q(\beta^*))^{\mathrm{T}}$, 其中

$$\varpi_{n,i}^s(\beta^*)=\big(Y_i-f(X_i,\beta^*)\big)\frac{\partial f(X_i,\beta^*)}{\partial\beta_s^*},\quad s=1,\cdots,q.$$

注意到, 当 $\beta^*=\beta$ 时, 有 $E[\varpi_{n,i}(\beta)]=0$. 此时, 可定义如下的经验对数似然比

$$\hat{l}_n(\beta^*)=-2\max\left\{\sum_{i=1}^n\log(np_i)\,\middle|\,p_i\geqslant 0,\sum_{i=1}^n p_i=1,\sum_{i=1}^n p_i\hat{\varpi}_{n,i}(\beta^*)=0\right\},$$

其中 $\{(\hat{Y}_i,\hat{X}_i),i=1,\cdots,n\}$ 是恢复数据, $\hat{\varpi}_{n,i}^s(\beta^*)=\big(\hat{Y}_i-f(\hat{X}_i,\beta^*)\big)\dfrac{\partial f(\hat{X}_i,\beta^*)}{\partial\beta_s^*}$, $s=1,\cdots,q$.

由 Lagrange 乘子法, 可将上述似然比 $\hat{l}_n(\beta^*)$ 重新表述为

$$\hat{l}_n(\beta^*)=2\sum_{i=1}^n\log\{1+\lambda^{\mathrm{T}}\hat{\varpi}_{n,i}(\beta^*)\},$$

其中 λ 是 Lagrange 乘子, 由下式所决定

$$\frac{1}{n}\sum_{i=1}^n\frac{\hat{\varpi}_{n,i}(\beta^*)}{1+\lambda^{\mathrm{T}}\hat{\varpi}_{n,i}(\beta^*)}=0.$$

定理 4.2.3 假设条件(C4.2.1)~条件(C4.2.9)成立, 当 $n\to\infty$ 时, 则有

$$\hat{l}_n(\beta)\xrightarrow{L}\chi_q^2.$$

基于定理 4.2.3, 可构造 β 的置信水平为 $1-\alpha$ 的经验似然置信域: $I_{\alpha,\mathrm{EL}}=\{\beta^*\in\mathbb{R}^q:\hat{l}_n(\beta^*)\leqslant c_\alpha\}$, 其中 c_α 是 χ_q^2 分布的上 α 分位数.

4.2.6 模拟研究

本节将用数值模拟来说明所提估计方法的完成情况. 从下面模型分别产生样本量 $n=200,300,400,500$ 和 600 的数据, 重复 500 次试验.

$$Y=\sin(\beta_1X_1)+(2+X_2)^{\beta_2}+\varepsilon, \tag{4.2.11}$$

其中 $\beta_1=1$, $\beta_2=0.5$. 模型误差 ε 来自正态分布 $N(0,0.5^2)$, 协变量 $(X_1,X_2)^{\mathrm{T}}$ 来自正态分布 $N(\mu_X,\Sigma)$, 其中 $\mu_X=(2,2)^{\mathrm{T}}$ 和

$$\Sigma=\begin{pmatrix}1&0.1\\0.1&0.25\end{pmatrix}.$$

失真函数 $\phi(\theta^{\mathrm{T}}U)=\dfrac{(2+\theta^{\mathrm{T}}U)^2}{C_Y}$, $\psi_1(\theta^{\mathrm{T}}U)=\dfrac{1.5+\theta^{\mathrm{T}}U}{C_{X_1}}$ 和 $\psi_2(\theta^{\mathrm{T}}U)=\dfrac{1+(\theta^{\mathrm{T}}U)^2}{C_{X_2}}$. 失真函数中的常数 (C_Y,C_{X_1},C_{X_2}) 需满足识别性条件: $E(\widetilde{Y})=E(Y)$, $E(\widetilde{X}_r)=E(X_r)$. 指标参数考虑下面两种情形.

情形 1 指标参数 θ 选取为 3 维向量 $\left(\dfrac{2}{\sqrt{29}},\dfrac{3}{\sqrt{29}},\dfrac{4}{\sqrt{29}}\right)^{\mathrm{T}}$, 混淆向量 U 取自正态分布 $N(\mu_U,\Sigma_U)$, 其均值向量和协方差矩阵分别是 $\mu_U=(4,5,6)^{\mathrm{T}}$ 和

$$\Sigma_U=\begin{pmatrix}1 & 0.4 & -0.2\\ 0.4 & 1 & 0.3\\ -0.2 & 0.3 & 1\end{pmatrix}.$$

指标参数 θ 和 U 的分布函数确定了 $C_Y=116.3869, C_{X_1}=10.2277$ 和 $C_{X_2}=78.4761$.

情形 2 指标参数 θ 选取为 6 维向量 $\left(\dfrac{1}{\sqrt{91}},\dfrac{2}{\sqrt{91}},\dfrac{3}{\sqrt{91}},\dfrac{4}{\sqrt{91}},\dfrac{5}{\sqrt{91}},\dfrac{6}{\sqrt{91}}\right)^{\mathrm{T}}$, 混淆向量 U 服从正态分布 $N(\mu_U,\Sigma_U)$, 其中 $\mu_U=(3,3,3,3,3,3)^{\mathrm{T}}$, $\Sigma_U=\left(0.5^{|i-j|}\right)_{1\leqslant i,j\leqslant 6}$. 此时, 常数 $C_Y=76.1915, C_{X_1}=8.1042$ 和 $C_{X_2}=46.7747$.

模拟结果报告在表 4.2.1~表 4.2.3 中, 表 4.2.3 报告了基于经验似然方法和渐近正态分布逼近方法所构造的参数 β 的置信水平为 95% 置信域的覆盖概率, 其中表中 "Bias" 表示参数估计的平均偏差, "SD" 表示参数估计的标准差, "EL" 表示经验似然方法, "NA" 表示基于渐近正态分布逼近的方法.

表 4.2.1 情形 1 参数估计的平均偏差 (Bias) 和标准差 (SD)

n		β_1	β_2	θ_1	θ_2	θ_3
$n=300$	Bias	0.0048	0.0249	0.0269	0.0317	−0.0361
	SD	0.0308	0.0668	0.0968	0.0746	0.0588
$n=400$	Bias	0.0050	0.0227	0.0192	0.0297	−0.0283
	SD	0.0247	0.0554	0.0869	0.0695	0.0575
$n=500$	Bias	0.0039	0.0149	0.0208	0.0282	−0.0260
	SD	0.0241	0.0533	0.0761	0.0646	0.0534
$n=600$	Bias	0.0040	0.0132	0.0154	0.0280	−0.0223
	SD	0.0206	0.0431	0.0695	0.0657	0.0520

表 4.2.2 情形 2 参数估计的平均偏差 (Bias) 和标准差 (SD)

n		β_1	β_2	θ_1	θ_2	θ_3	θ_4	θ_5	θ_6
$n=300$	Bias	−0.0017	−0.0016	−0.0099	0.0194	0.0088	−0.0045	−0.0015	−0.0271
	SD	0.0264	0.0220	0.0939	0.0711	0.0679	0.0564	0.0522	0.0480
$n=400$	Bias	0.0008	−0.0007	−0.0014	0.0126	0.0062	0.0064	−0.0029	−0.0264
	SD	0.0219	0.0186	0.0767	0.0637	0.0582	0.0528	0.0520	0.0473

续表

n		β_1	β_2	θ_1	θ_2	θ_3	θ_4	θ_5	θ_6
$n=500$	Bias	−0.0001	−0.0009	−0.0055	0.0124	0.0071	0.0021	−0.0059	−0.0177
	SD	0.0203	0.0176	0.0691	0.0573	0.0552	0.0501	0.0470	0.0424
$n=600$	Bias	0.0040	−0.0004	−0.0011	0.0134	0.0023	0.0039	−0.0082	−0.0126
	SD	0.0175	0.0160	0.0664	0.0531	0.0480	0.0477	0.0424	0.0389

表 4.2.3 参数 $(\beta_1,\beta_2)^{\mathrm{T}}$ 的置信水平为 95%置信域的覆盖概率

情形 1				
方法	$n=300$	$n=400$	$n=500$	$n=600$
EL	91.2%	93.4%	94.2%	94.6%
NA	85.8%	89.8%	92.2%	93.9%
情形 2				
方法	$n=300$	$n=400$	$n=500$	$n=600$
EL	92.0%	93.1%	94.0%	94.5%
NA	90.7%	90.1%	93.5%	94.2%

从表 4.2.1~表 4.2.3 中结果可以看出：

(1) 随着样本量的增大, $(\beta_1,\beta_2)^{\mathrm{T}}$ 的估计值越来越接近于真值 $(1,0.5)^{\mathrm{T}}$, 并且指标参数 θ 的估计 $\hat{\theta}$ 也很接近于真值;

(2) 从表 4.2.3 可以看出, 经验似然方法所构造的置信域的覆盖概率更靠近真实的置信水平 95%.

图 4.2.1 和图 4.2.2 模拟了一次样本量为 400 时, $(\beta_1,\beta_2)^{\mathrm{T}}$ 在两种情形下置信水平为 95%的置信域. 从图 4.2.1 和图 4.2.2 可以看出, 经验似然方法所构造的置信域要比基于渐近正态分布逼近所构造的置信域要小. 表明经验似然方法在置信域构造方面要优于渐近正态分布逼近方法.

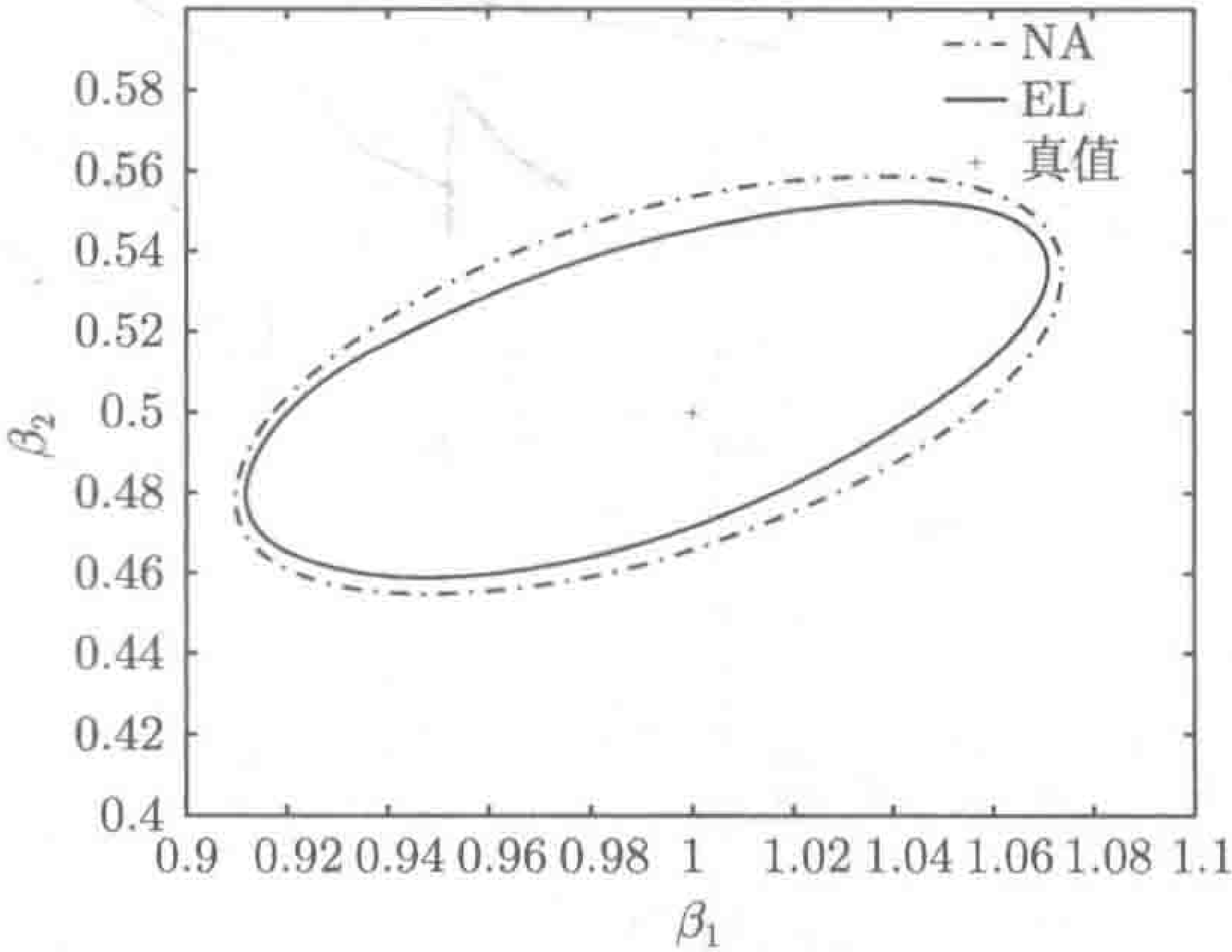


图 4.2.1 情形 1 时参数 $(\beta_1,\beta_2)^{\mathrm{T}}$ 置信水平为 95%的置信域. 实线表示用经验似然方法得到的置信域, 虚线表示基于渐近正态分布逼近所得到的置信域, 加号表示参数 $(\beta_1,\beta_2)^{\mathrm{T}}$ 的真值

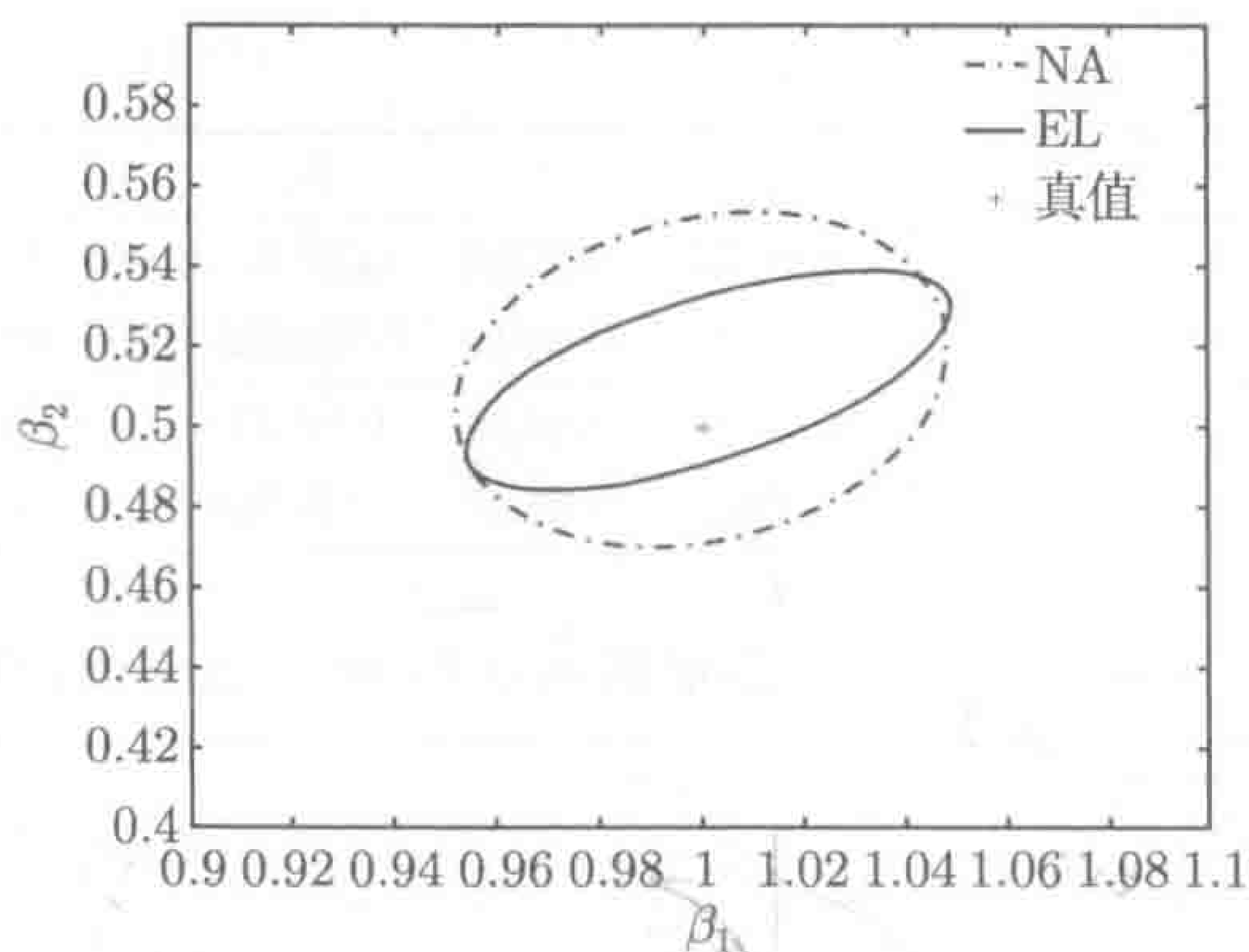


图 4.2.2 情形 2 时参数 $(\beta_1, \beta_2)^{\mathrm{T}}$ 置信水平为 95%的置信域. 实线表示用经验似然方法得到的置信域, 虚线表示基于渐近正态分布逼近所得到的置信域, 加号表示参数 $(\beta_1, \beta_2)^{\mathrm{T}}$ 的真值

4.2.7 实例分析

本节分析一组皮马印第安人糖尿病实际数据, 说明所提方法的实用性. Nguyen 和Sentürk (2008) 建议使用体重指标 (BMI) 和三头肌皮褶厚度 (SFT) 作为混淆变量. 在正常情况下, 葡萄糖的吸收机制会被高水平的血清胰岛素所破坏, 尤其是对 II- 型的糖尿病患者而言. 因此, Zhang 等 (2012a) 分析这组数据中的 380 个样本时, 关心的问题是血糖浓度 (GLU) 和 2 小时血清胰岛素 (SER) 之间的关系.

对该组数据进行分析时, 失真函数中单指标参数 θ 估计为 $\hat{\theta} = (0.7579, 0.6524)^{\mathrm{T}}$. 因此, 单指标混淆变量 $\theta^{\mathrm{T}}U$ 的估计为 0.7579SFT+0.6524BMI. 再根据式 (4.2.9) 可得到失真函数 $\phi(\cdot)$ 和 $\psi(\cdot)$ 的估计, 并在图 4.2.3 中分别给出各自的估计曲线和 95% 的置信带. 图中这两个曲线都显示 $\phi(u)$ 和 $\psi(u)$ 是非线性函数, 也表明单指标混淆变量 $\theta^{\mathrm{T}}U$ 确实对血糖浓度与 2 小时血清胰岛素有影响.

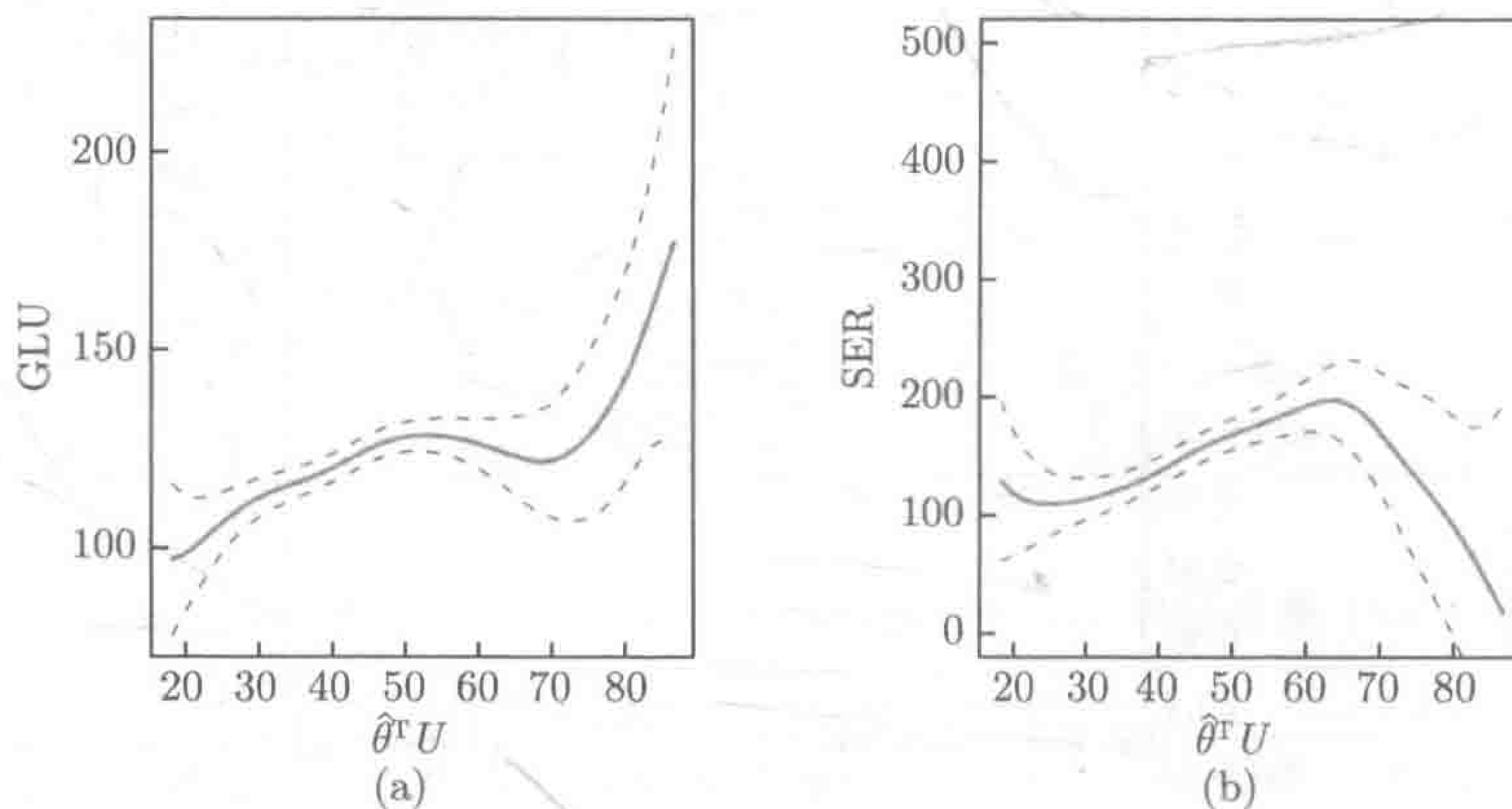


图 4.2.3 失真函数 $\phi(\cdot)$ 的局部线性光滑估计, 以及置信水平为 95%的置信带 ((a) 图), 失真函数 $\psi(\cdot)$ 的局部线性光滑估计, 以及置信水平为 95%的置信带 ((b) 图)

下面用估计的单指标混淆变量 0.7579SFT+0.6524BMI 和式 (4.2.9) 来得到该组数据中变量 GLU 和变量 SER 的估计, 分别记为 $\widehat{\text{GLU}}$, $\widehat{\text{SER}}$. 并用 Fan 和 Gijbels(1996) 提出的局部多项式方法建立 $\widehat{\text{GLU}}$ 和 $\widehat{\text{SER}}$ 之间的非参数曲线拟合. 这条估计出来的曲线和 95%的置信带在图 4.2.4 中展示. 同时, 在图 4.2.4 中建立 $\widehat{\text{GLU}}$ 和 $\widehat{\text{SER}}$ 的线性回归关系. 这条拟合的直线并不在非参数曲线的 95%的置信带内, 说明该数据中的 GLU 和 SER 之间存在非线性的关系. 因此, Zhang 等 (2012a) 采用由 Lee 和 Kwon (2004) 引入的非线性模型来研究血糖浓度与 2 小时血清胰岛素的关系, 即考虑

$$\text{GLU} = f(\text{SER}, \beta) + \varepsilon = \frac{\beta_1 + \beta_2 \text{SER}}{\beta_3 + \text{SER}} + \varepsilon. \tag{4.2.12}$$

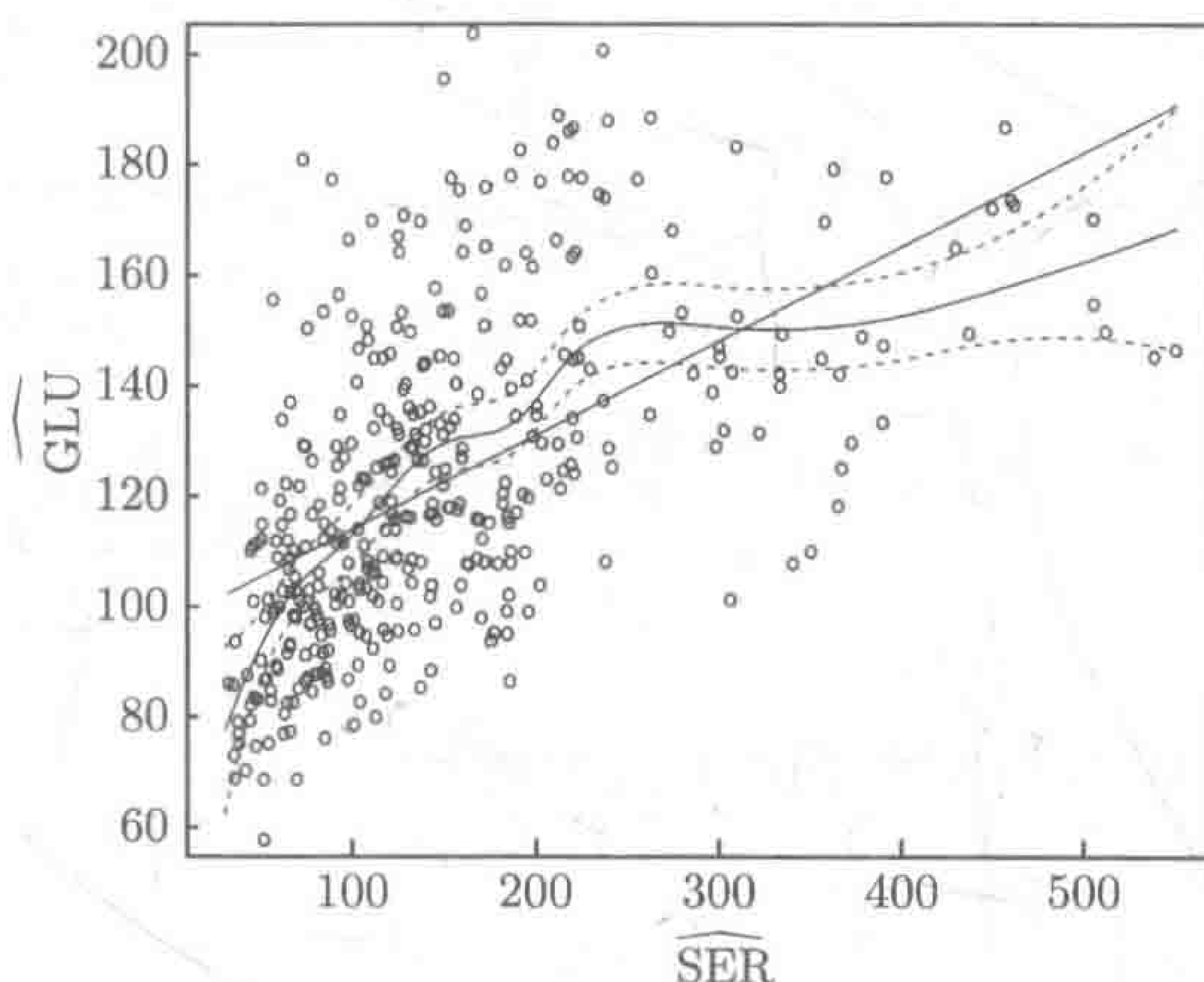


图 4.2.4　实线表示 $\widehat{\text{GLU}}$ 和 $\widehat{\text{SER}}$ 之间的非参数曲线的局部线性光滑估计, 虚线表示置信水平为 95%的置信带, 直线表示线性拟合结果

利用非线性最小二乘方法, 可得 $(\hat{\beta}_1, \hat{\beta}_2, \hat{\beta}_3)^{\text{T}} = (4441.10, 182.02, 96.19)^{\text{T}}$. 此外, 参数 $(\beta_1, \beta_2, \beta_3)^{\text{T}}$ 每个分量基于定理 4.2.2 的渐近正态分布结果所构造的 95%的置信区间分别为 $(-2180, 11063)$, (161.57, 202.47), $(20.15, 172.25)$, 而经验似然方法所构造的置信区间分别为 $(4000.6, 4908.7)$, $(178.01, 186.19)$ 和 $(91.44, 101.03)$. 从上述估计也可以看出, 当 SER 取值较大的时候, 血糖浓度应该在 182 附近波动. 值得注意的是, 基于定理 4.2.2 的渐近正态分布结果所构造的置信区间的长度要远远大于经验似然方法所构造置信区间的长度. 对参数 β_1 而言, 渐近正态分布结果所构造的置信区间包含 0, 而经验似然方法构造的置信区间不包含 0. 这两种不同的结果也导致了有争议的问题, 即 β_1 是否显著. 然而, 通过 4.2.6 节的模拟研究, 这里更倾向于选择经验似然方法的结论: 即参数 β_1 的真值非零.

下面用模型检验的思路来考察上述使用的模型 (4.2.12) 是否适合于这组数据.

Zhang 等 (2012a) 使用了 Stute 等 (1998b) 提出的经验过程检验统计量, 得到的统计量值为 0.0497, p 值为 0.9790. 该检验结果表明模型 (4.2.12) 可用于分析这组数据. 最后, 图 4.2.5 展示了模型 (4.2.12) 的拟合曲线和置信水平为 95%的置信带.

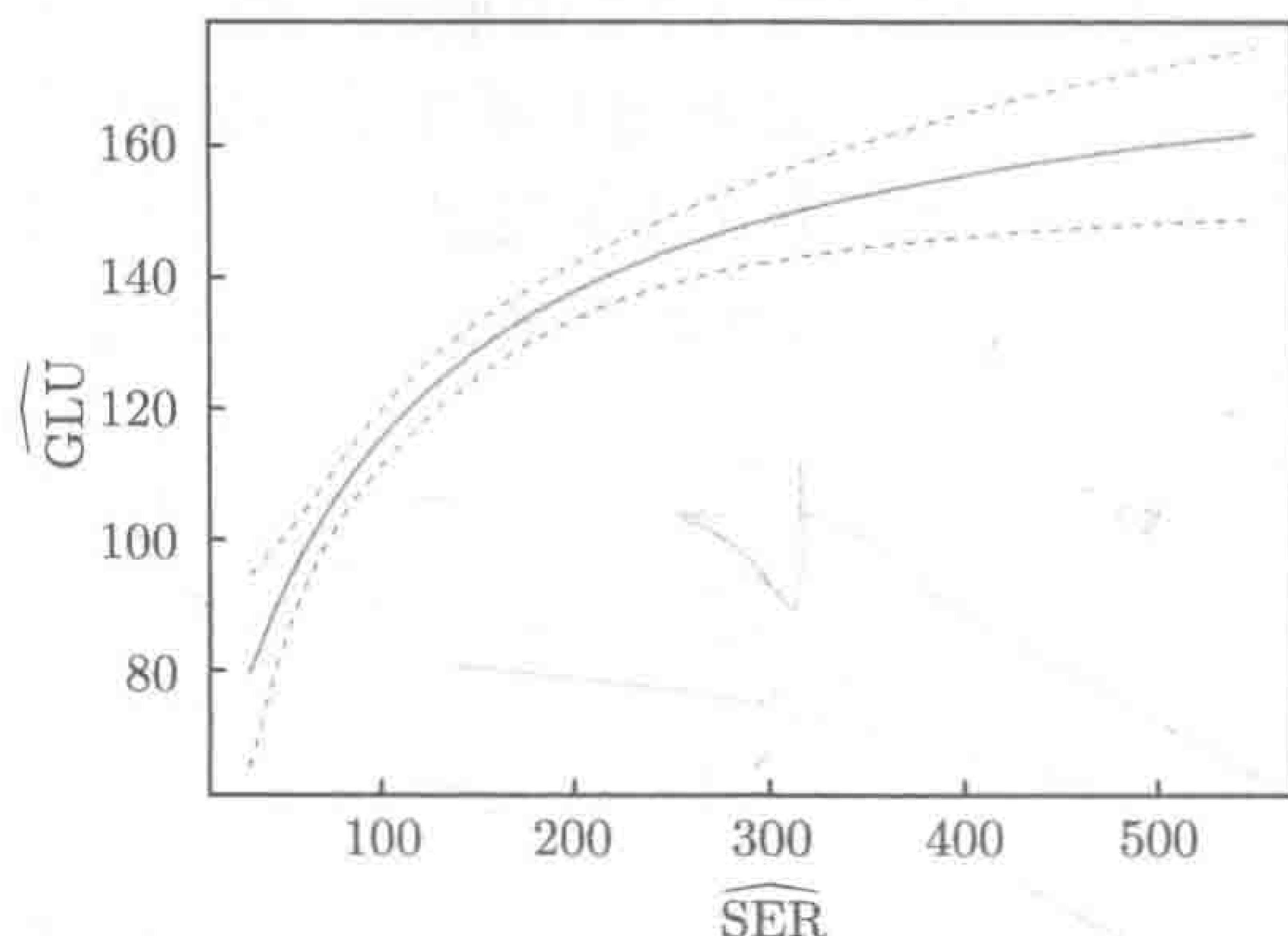


图 4.2.5　非线性函数 $f(\cdot,\cdot)$ 的曲线拟合和置信水平为 95%的置信带

4.3　可加协变量调整非线性模型

4.3.1　模型介绍

本节考虑下面的可加协变量调整非线性模型:

$$\begin{cases} Y = f(X,\beta) + \varepsilon, \\ \widetilde{Y} = \phi(U)Y, \\ \widetilde{X}_r = \psi_r(U)X_r, \quad r = 1,\cdots,p, \end{cases} \tag{4.3.1}$$

其中 $f(X,\beta)$ 是函数形式已知的非线性函数, Y 是不可观测的响应变量, β 是未知的 q 维参数向量, $X = (X_1,\cdots,X_p)^{\mathrm{T}}$ 是 p 维的不可观测的连续型协变量, $U = (U_1,\cdots,U_d)^{\mathrm{T}}$ 是 d 维可观测的混淆向量, $\phi(U)$ 和 $\psi_r(U)$ 是混淆向量 U 的未知的失真函数, $\widetilde{Y}$ 和 $\widetilde{X}_r$ 分别是受失真函数 $\phi(U)$ 和 $\psi_r(U)$, 以及混淆变量 U 污染以后得到的可观测的响应变量和协变量, $r = 1,\cdots,p$.

在多元混淆向量 $U = (U_1,\cdots,U_d)^{\mathrm{T}}$ 的这一情形下, 为了避免 "维数祸根" 问题, Zhang 等 (2014b) 用可加模型来构建失真函数, 即考虑

$$\begin{cases} \phi(U) = \alpha_0 + \phi_1(U_1) + \cdots + \phi_d(U_d) =: \alpha_0 + \phi^o(U), \\ \psi_r(U) = \alpha_r + \psi_{r1}(U_1) + \cdots + \psi_{rd}(U_d) =: \alpha_r + \psi_r^o(U), \quad r = 1,\cdots,p. \end{cases} \tag{4.3.2}$$

在模型 (4.3.2) 中, $\phi_s(U_s)$ 和 $\psi_{rs}(U_s)$ 是未知的失真函数. 关于可加模型研究

的一些介绍, 读者可参阅 Hastie 和 Tibshirani (1990), Stone (1985, 1986), Huang (2003), Li 和 Ruppert (2008), Zhang 和 Liang (2011) 和 Wang 等 (2011).

为了介绍估计程序, 需引入下面关于模型 (4.3.1) 和模型 (4.3.2) 基本的假设条件:

(1) (X, U, ε) 是相互独立的, 且 (Y, U) 也是相互独立的;

(2) $E(\varepsilon)=0, \mathrm{Var}(\varepsilon)=\sigma^2<\infty$;

(3) $E[\phi(U)]=1, E[\psi_r(U)]=1, r=1,\cdots,p$;

(4) $E[\phi_s(U_s)]=0$, $E[\psi_{rs}(U_s)]=0$, $s=1,\cdots,d$, $r=1,\cdots,p$.

假设条件 (3) 是失真函数的识别性条件, 假设条件 (4) 是可加模型的识别性条件. 这两个条件确保了可加失真模型中的参数 (α_0, α_r) 满足 $\alpha_0=1, \alpha_r=1$, $r=1,\cdots,p$.

4.3.2 估计方法和渐近性质

不失一般性, 假设混淆变量 U_s 在闭区间 $[0,1]$ 上取值, $s=1,\cdots,d$. 在实际中, 可将混淆变量 U_s 进行一些单调变换, 使得变换之后的变量取值在闭区间 $[0,1]$ 上. 下面采用样条方法来估计失真函数 $\phi^o(\cdot)$ 和 $\psi_r^o(\cdot), r=1,\cdots,p$.

记 $\mathcal{S}_n$ 是定义在闭区间 $[0,1]$ 上样条光滑度 ϱ 满足 $\varrho\geqslant 1$ 的多项式样条集. 该集合中, 多项式样条的节点设定为 $k_{-\varrho}=\cdots=k_{-1}=k_0=0<k_1<\cdots<k_J<1=k_{J+1}=\cdots=k_{J+\varrho+1}$, 其中 $J=J_n$ 是随着样本容量 n 的增大而增大的递增序列. 集合 $\mathcal{S}_n$ 中的每个样条函数 η 满足下面两个条件:

条件 1 η 在每个子区间 $[k_t, k_{t+1}), t=0,1,\cdots,J_n-1$ 和最后一个区间 $[k_{J_n}, 1]$ 上都是 ϱ 阶多项式;

条件 2 η 在区间 $[0,1]$ 上是 $\varrho-1$ 阶连续可微函数, 其中 $\varrho\geqslant 2$.

定义 $h=\dfrac{1}{J_n+1}$ 是两个等距节点之间的长度, 本节采用等距节点的多项式样条逼近来对失真函数进行估计. 取 $b_{j,s}(u_s)$ 是样条光滑度为 ϱ 的 B 样条基函数. 对每个失真函数 $\phi(u)$ 和 $\psi_s(u)$ 中的可加部分 $\phi_s(u_s)$ 和 $\psi_{rs}(u_s)$, 这里采用 B 样条逼近的方法, 即

$$\begin{aligned}\phi_s(u_s)&\approx\gamma_{0,-\varrho,s}b_{-\varrho,s}(u_s)+\gamma_{0,-\varrho+1,s}b_{-\varrho+1,s}(u_s)+\cdots+\gamma_{0,J_n,s}b_{J_n,s}(u_s),\\ \psi_{rs}(u_s)&\approx\gamma_{r,-\varrho,s}b_{-\varrho,s}(u_s)+\gamma_{r,-\varrho+1,s}b_{-\varrho+1,s}(u_s)+\cdots+\gamma_{r,J_n,s}b_{J_n,s}(u_s),\end{aligned}$$

其中 $\gamma_{0,j,s}$ 和 $\gamma_{r,j,s}$ 是待估系数, $j=-\varrho,-\varrho+1,\cdots,J_n$, $s=1,\cdots,d$, $r=1,\cdots,p$.

因此, 可以构造下面的目标函数

$$
\begin{cases}
\sum_{i=1}^{n}\left\{\widetilde{Y}_i\Big/\overline{\widetilde{Y}}-\alpha_0-\sum_{s=1}^{d}\sum_{j=-\varrho}^{J_n}\gamma_{0,j,s}b(U_{is})\right\}^2, \\
\sum_{i=1}^{n}\left\{\widetilde{X}_{ir}\Big/\overline{\widetilde{X}}_r-\alpha_r-\sum_{s=1}^{d}\sum_{j=-\varrho}^{J_n}\gamma_{r,j,s}b(U_{is})\right\}^2, \quad r=1,\cdots,p,
\end{cases}
\tag{4.3.3}
$$

其中 $\overline{\widetilde{Y}}=\frac{1}{n}\sum_{i=1}^{n}\widetilde{Y}_i$, $\overline{\widetilde{X}}_{ri}=\frac{1}{n}\sum_{i=1}^{n}\widetilde{X}_{ir}$. 记 $\hat{\alpha}_0$, $\hat{\alpha}_r$, $\hat{\gamma}_{0,j,s}$ 和 $\hat{\gamma}_{r,j,s}$ 是极小化目标函数 (4.3.3) 得到的估计, $r=1,\cdots,p$, $j=-\varrho,-\varrho+1,\cdots,J_n$, $s=1,\cdots,d$. 注意到可加函数需满足的识别性条件 $E(\phi_r(U_s))=0$ 和 $E(\psi_{rs}(U_s))=0$, $s=1,\cdots,d$, $r=1,\cdots,p$. 因此, 最终失真函数的估计可定义为

$$
\begin{aligned}
\hat{\phi}(u)&=\hat{\alpha}_0+\sum_{s=1}^{d}\hat{\phi}_s^*(u_s),\\
\hat{\psi}_r(u)&=\hat{\alpha}_r+\sum_{s=1}^{d}\hat{\psi}_{rs}^*(u_s), \quad s=1,\cdots,d, \quad r=1,\cdots,p,
\end{aligned}
$$

其中

$$
\hat{\phi}_s^*(u_s)=\sum_{j=-\varrho}^{J_n}\hat{\gamma}_{0,j,s}b_{j,s}(u_s)-\frac{1}{n}\sum_{i=1}^{n}\sum_{j=-\varrho}^{J_n}\hat{\gamma}_{0,j,s}b_{j,s}(U_{is}), \tag{4.3.4}
$$

$$
\hat{\psi}_{rs}^*(u_s)=\sum_{j=-\varrho}^{J_n}\hat{\gamma}_{r,j,s}b_{j,s}(u_s)-\frac{1}{n}\sum_{i=1}^{n}\sum_{j=-\varrho}^{J_n}\hat{\gamma}_{r,j,s}b_{j,s}(U_{is}). \tag{4.3.5}
$$

进一步, 利用关系式 $Y=\frac{\widetilde{Y}}{\phi(U)}$ 和 $X_r=\frac{\widetilde{X}_r}{\psi_r(U)}$, 可使用 Cui 等 (2009) 提出的直接估计方法将不可观测的响应变量 Y 和协变量 X 估计出来, 定义为

$$
\hat{Y}_i=\frac{\widetilde{Y}_i}{\hat{\phi}(U_i)}, \quad \hat{X}_{ir}=\frac{\widetilde{X}_{ir}}{\hat{\psi}_r(U_i)}, \quad r=1,\cdots,p, \tag{4.3.6}
$$

其中

$$
\begin{aligned}
\hat{\phi}(U_i)&=\hat{\alpha}_0+\sum_{s=1}^{d}\left\{\sum_{j=-\varrho}^{J_n}\hat{\gamma}_{0,j,s}b_{j,s}(U_{is})-\frac{1}{n}\sum_{i=1}^{n}\sum_{j=-\varrho}^{J_n}\hat{\gamma}_{0,j,s}b_{j,s}(U_{is})\right\},\\
\hat{\psi}_r(U_i)&=\hat{\alpha}_r+\sum_{s=1}^{d}\left\{\sum_{j=-\varrho}^{J_n}\hat{\gamma}_{r,j,s}b_{j,s}(U_{is})-\frac{1}{n}\sum_{i=1}^{n}\sum_{j=-\varrho}^{J_n}\hat{\gamma}_{r,j,s}b_{j,s}(U_{is})\right\}.
\end{aligned}
$$

基于上面的讨论, 这时可得到不可观测变量 (Y_i, X_i) 的恢复数据 $(\hat{Y}_i, \hat{X}_i), i=1,\cdots,n$. 基于恢复数据集 $\{(\hat{Y}_i, \hat{X}_i), i=1,\cdots,n\}$ 和非线性最小二乘方法, 参数 β 的估计可通过求解下面定义的估计方程得到:

$$\sum_{i=1}^{n}[\hat{Y}_i - f(\hat{X}_i,\beta)]\frac{\partial f(\hat{X}_i,\beta)}{\partial\beta_r} = 0, \quad r=1,\cdots,p, \tag{4.3.7}$$

其中 $\dfrac{\partial f(\cdot,\beta)}{\partial\beta_r}$ 是已知非线性函数 $f(\cdot,\beta)$ 关于第 r 个分量 β_r 的偏导数. 当估计方程 (4.3.7) 没有显式表达式时, 可使用 Newton-Raphson 迭代算法进行求解. 把解估计方程 (4.3.7) 得到的估计称为非线性最小二乘估计, 记为 $\hat{\beta}$.

为了给出参数估计 $\hat{\beta}$ 的渐近性质, 下面列出所需要的正则条件.

定义 $\|g\|_\infty = \sup\limits_t|g(t)|$, 表示函数 $g(\cdot)$ 在区间 $[0,1]$ 上的模. 记 ν 是一个正整数, ι 是取值在 $(0,1]$ 上的一个实数, 并满足 $v=\nu+\iota>2$. 集合 $\mathcal{F}$ 中的每一个函数 $g(\cdot)$ 定义在区间 $[0,1]$ 上, $g(\cdot)$ 的 ν 阶导数 $g^{(\nu)}(\cdot)$ 存在, 并满足下面的 Lipschitz 条件:

$$|g^{(\nu)}(t_1) - g^{(\nu)}(t_2)| \leqslant C_o|t_1-t_2|^\iota, \quad 0\leqslant t_1, t_2\leqslant 1,$$

其中 C_o 是一个正常数.

(C4.3.1) 失真函数 $\phi(t)$, $\psi_r(t)$, $r=1,\cdots,p$, 在 $t\in\mathcal{C}$ 上不恒为常数且不存在零点, 并有 $\phi_s\in\mathcal{F}$, $\psi_{rs}\in\mathcal{F}$, $r=1,\cdots,p$, $s=1,\cdots,d$.

(C4.3.2) 模型误差 ε 满足 $E\varepsilon^4<\infty$.

(C4.3.3) $E(Y)\neq 0$ 和 $E(X_r)\neq 0$, $r=1,\cdots,p$.

(C4.3.4) 当 $l_1,l_2,l_3,l_4=0,1,2$, $l_1+l_2+l_3+l_4\leqslant 3$, $1\leqslant s_1,s_2\leqslant p$ 和 $1\leqslant t_1,t_2\leqslant q$, 偏导数

$$\frac{\partial^{l_1+l_2+l_3+l_4}f(X,\beta)}{\partial^{l_1}\beta_{t_1}\partial^{l_2}\beta_{t_2}\partial^{l_3}X_{s_1}\partial^{l_4}X_{s_2}}$$

存在, 并存在正数 C_1 使得下面的不等式

$$\left|\frac{\partial^{l_1+l_2+l_3+l_4}f(X,\beta)}{\partial^{l_1}\beta_{t_1}\partial^{l_2}\beta_{t_2}\partial^{l_3}X_{s_1}\partial^{l_4}X_{s_2}}\right|\leqslant C_1$$

在 $l_3+l_4\geqslant 1$ 成立. 同时存在正数 C_2 使得下面的不等式

$$E\left\{\sup_\beta\left|\frac{\partial^{l_1+l_2+l_3+l_4}f(X,\beta)}{\partial^{l_1}\beta_{t_1}\partial^{l_2}\beta_{t_2}\partial^{l_3}X_{s_1}\partial^{l_4}X_{s_2}}\right|\right\}\leqslant C_2$$

在 $1\leqslant l_1+l_2\leqslant 2$ 和 $l_3+l_4=0$ 时成立.

(C4.3.5) 参数真值 β 满足: $\beta = \arg\min_{\beta^* \in \mathbb{R}^q} E\big[f(X,\beta) - f(X,\beta^*)\big]^2$.

(C4.3.6) 样条节点数 J_n 满足当 $n \to \infty$ 时, $\dfrac{J_n}{n^{1/3}} \to 0$, $\dfrac{J_n}{n^{1/(2v)}} \to \infty$.

定义记号:

$$\Gamma = E\{f_\beta^{'\otimes 2}(X,\beta)\}, \qquad \mathcal{I} = \left(E\left[\frac{Yf_1'(X,\beta)}{\phi(U)}\right], \cdots, E\left[\frac{Yf_q'(X,\beta)}{\phi(U)}\right]\right)^{\mathrm{T}},$$

其中 $f_\beta'(X,\beta) = (f_1'(X,\beta), \cdots, f_q'(X,\beta))^{\mathrm{T}}$, $f_l'(X,\beta) = \dfrac{\partial f(X,\beta)}{\partial \beta_l}$, $l = 1, \cdots, q$. 定义 $\mathcal{C} = (c_{r,l}(\beta))_{1\leqslant r\leqslant q, 1\leqslant l\leqslant p}$, $c_{r,l}(\beta) = E\left[\dfrac{X_l f_{X_l}'(X,\beta) f_r'(X,\beta)}{\psi_l(U)}\right]$, $f_{X_l}'(X,\beta) = \dfrac{\partial f(X,\beta)}{\partial X_l}$, $l = 1, \cdots, p$. 进一步, 定义 $D_\psi = (\psi_1(U) - 1, \cdots, \psi_p(U) - 1)^{\mathrm{T}}$ 和 $\Sigma_\psi = E(D_\psi^{\otimes 2})$.

定理 4.3.1 假设条件(C4.3.1)~条件(C4.3.6)成立, 当 $n \to \infty$ 时, 则有

$$\sqrt{n}(\hat{\beta} - \beta) \xrightarrow{L} N(0, \Gamma^{-1}\Sigma\Gamma^{-1}),$$

其中 $\Sigma = \Gamma\sigma^2 + \mathrm{Var}(\phi(U))\mathcal{I}^{\otimes 2} + \mathcal{C}\Sigma_\psi\mathcal{C}^{\mathrm{T}} - \mathcal{I}E\{\phi(U)D_\psi^{\mathrm{T}}\}\mathcal{C}^{\mathrm{T}} - \mathcal{C}E\{\phi(U)D_\psi\}\mathcal{I}^{\mathrm{T}}$.

4.3.3 经验似然置信域

类似于 4.2.5 节, 本节也可以使用经验似然方法构造参数 β 的经验似然置信域. 定义下面的经验对数似然比

$$\hat{l}_n(\beta^*) = -2\max\left\{\sum_{i=1}^n \log(np_i) \,\middle|\, p_i \geqslant 0, \sum_{i=1}^n p_i = 1, \sum_{i=1}^n p_i \hat{\varpi}_{n,i}(\beta^*) = 0\right\},$$

其中 $\hat{\varpi}_{n,i}^s(\beta^*) = \big(\hat{Y}_i - f(\hat{X}_i,\beta^*)\big)\dfrac{\partial f(\hat{X}_i,\beta^*)}{\partial \beta_s^*}$ 是辅助随机变量, $s = 1, \cdots, q$.

由 Lagrange 乘子法, 经验对数似然比 $\hat{l}_n(\beta^*)$ 能够重新写为

$$\hat{l}_n(\beta^*) = 2\sum_{i=1}^n \log\{1 + \lambda^{\mathrm{T}}\hat{\varpi}_{n,i}(\beta^*)\},$$

其中 λ 是 Lagrange 乘子, 由下式决定

$$\frac{1}{n}\sum_{i=1}^n \frac{\hat{\varpi}_{n,i}(\beta^*)}{1 + \lambda^{\mathrm{T}}\hat{\varpi}_{n,i}(\beta^*)} = 0.$$

定理 4.3.2 假设条件(C4.3.1)~条件(C4.3.6) 成立, 当 $n \to \infty$ 时, 且 β 是参数真值时, 则有

$$\hat{l}_n(\beta) \xrightarrow{L} \chi_q^2.$$

基于定理 4.2.3, 能够构造 β 置信水平为 $1 - \alpha$ 的经验似然置信域为 $I_{\alpha,\mathrm{EL}} = \{\beta^* \in \mathbb{R}^q : \hat{l}_n(\beta^*) \leqslant c_\alpha\}$, 其中 c_α 是 χ_q^2 分布的上 α 分位数.

4.3.4 模型检验

模型检验考虑响应变量 Y 与自变量 X 之间是否可以用事先给定的模型进行数据拟合. 在协变量调整数据背景下, 考虑的问题是不可观测的响应变量 Y 与不可观测的协变量 X 之间的回归模型 $E(Y|X)=g(X)$ 是否可以用事先给定的参数模型进行拟合. 若记 $\mathcal{M}=\{g(\cdot,\beta):\beta\in\Theta\subset\mathbb{R}^q\}$ 是一个给定的参数函数集合, β 是未知的 q 维真实参数, 那么模型检验可归结为如下假设检验问题.

原假设 H_0 : 存在某个参数向量 β与某个参数函数$g(\cdot,\cdot)\in\mathcal{M}$, 有

$$g(X)=g(X,\beta),$$

备择假设 H_1 : 对任意的参数向量 β^* 与任意的参数函数 $g(\cdot,\cdot)\in\mathcal{M}$, 都有

$$g(X)\neq g(X,\beta^*).$$

与 2.8 节基于残差构造经验过程来检验线性模型的方法不同, 本节采用一种基于协变量构造经验过程检验统计量的模型检验方法. 关于该经验过程检验方法的讨论, 见 Stute 等 (1998a). 基于恢复数据集 $\{(\hat{Y}_i,\hat{X}_i),i=1,\cdots,n\}$, 定义经验过程统计量

$$\mathcal{T}_n(x_1,\cdots,x_p)=n^{-1/2}\sum_{i=1}^{n}I\{\hat{X}_{i1}\leqslant x_1,\cdots,\hat{X}_{ip}\leqslant x_p\}\hat{\varepsilon}_i, \tag{4.3.8}$$

其中 $\hat{\varepsilon}_i=\hat{Y}_i-f(\hat{X}_i,\hat{\beta})$, $1\leqslant i\leqslant n$, $I\{\cdot\}$ 是示性函数. 基于由式 (4.3.8) 定义的经验过程统计量 $\mathcal{T}_n(x_1,\cdots,x_p)$, Zhang 等 (2014b) 构造了下面的 Cramér-von Mises (CvM) 统计量

$$\mathcal{W}_n=\frac{1}{n}\sum_{i=1}^{n}\mathcal{T}_n^2(\hat{X}_{i1},\cdots,\hat{X}_{ip}). \tag{4.3.9}$$

直观上讲, 原假设 H_0 成立时, 统计量 $\mathcal{W}_n$ 的取值偏小; 原假设 H_0 不成立时, 统计量 $\mathcal{W}_n$ 的取值就会偏大. 因此当统计量 $\mathcal{W}_n$ 的值比较大的时候, 就应该拒绝原假设 H_0. 为了找到拒绝原假设 H_0 的临界值, 采用 wild-bootstrap 算法来近似检验统计量 $\mathcal{W}_n$ 在原假设 H_0 下的渐近分布. 具体算法如下.

步骤 1 基于恢复数据集 $\{(\hat{Y}_i,\hat{X}_i),i=1,\cdots,n\}$ 和非线性最小二乘估计 $\hat{\beta}$, 计算残差 $\hat{\varepsilon}_i=\hat{Y}_i-f(\hat{X}_i,\hat{\beta})$, 进一步计算检验统计量 $\mathcal{W}_n$.

步骤 2 独立产生 N 个 Bernoulli 分布的随机数 ν_k, $k=1,\cdots,N$, 满足

$$P\left(\nu_k=\frac{1-\sqrt{5}}{2}\right)=\frac{5+\sqrt{5}}{10},\quad P\left(\nu_k=\frac{1+\sqrt{5}}{2}\right)=\frac{5-\sqrt{5}}{10}.$$

利用此随机数产生 N 个新的响应变量 $\hat{Y}_{ik}^*$, 即

$$\hat{Y}_{ik}^* = f(\hat{X}_i, \hat{\beta}) + \hat{\varepsilon}_i \nu_k.$$

步骤 3 基于新的响应变量 $\hat{Y}_{ik}^*$ 和非线性最小二乘方法, 求解下面定义的 q 个估计方程, 可得到 β 的估计 $\hat{\beta}^{[k]}$. 估计方程定义如下

$$\sum_{i=1}^{n}[\hat{Y}_{ik}^* - f(\hat{X}_i, \beta)]\frac{\partial f(\hat{X}_i, \beta)}{\partial \beta_r} = 0, \quad r = 1, \cdots, p, \quad k = 1, \cdots, N.$$

步骤 4 计算残差 $\hat{\varepsilon}_{ik}^* = \hat{Y}_{ik}^* - f(\hat{X}_i, \hat{\beta}^{[k]})$, 以及 N 个检验统计量 $\mathcal{T}_{n,k}$ 和 $\mathcal{W}_{n,k}$ 分别为

$$\mathcal{T}_{n,k}(x_1, \cdots, x_p) = n^{-1/2}\sum_{i=1}^{n} I\{\hat{X}_{i1} \leqslant x_1, \cdots, \hat{X}_{ip} \leqslant x_p\}\hat{\varepsilon}_{ik}^*$$

和

$$\mathcal{W}_{n,k} = \mathcal{T}_{n,k}^2(\hat{X}_{i1}, \cdots, \hat{X}_{ip}), \quad k = 1, \cdots, N.$$

步骤 5 使用 $\{\mathcal{W}_{n,k}, k = 1, \cdots, N\}$ 的经验分布函数来逼近检验统计量 $\mathcal{W}_n$ 的渐近分布, 进而计算 p 值为

$$\hat{p} = \frac{1}{N}\sum_{k=1}^{N} I\left\{\mathcal{W}_{n,k} \geqslant \mathcal{W}_n\right\}.$$

4.3.5 模拟研究

本节将用数值模拟研究来说明所提估计方法和模型检验方法的完成情况, 例 4.3.1 说明估计方法的完成情况, 例 4.3.2 说明所提出的 wild-bootstrap 逼近的模型检验方法在有限样本下的完成情况.

例 4.3.1 从下面的非线性模型:

$$Y = \beta_1 + X_1(1 + X_2)^{\beta_2} + \varepsilon \tag{4.3.10}$$

分别产生样本量 $n = 200, 400$ 和 600 的 500 个数据集, 其中 $\beta_1 = 1$, $\beta_2 = -0.5$, 模型误差 ε 来自正态分布 $N(0, 0.5^2)$, 协变量 $(X_1, X_2)^{\mathrm{T}}$ 来自正态分布 $N(\mu_X, \Sigma)$, 其中 $\mu_X = (2, 10)^{\mathrm{T}}$ 和 $\Sigma = \left(0.5^{|i-j|}\right)_{1\leqslant i,j\leqslant 2}$. 失真函数取为

$$\begin{aligned}
\phi(U) &= 1 + 0.2\cos(2\pi U_1) + 0.3(U_2^4 - 0.2),\\
\psi_1(U) &= 1 + 0.3\cos(2\pi U_1) + 0.3(U_2^3 - 0.25),\\
\psi_2(U) &= 1 + 0.2(U_1^2 - 1/3) + 0.3\left(\exp(-6.5U_2) - 0.1536\right),
\end{aligned}$$

其中 $(U_1,U_2)^{\mathrm{T}}$ 相互独立且都来自于 $[0,1]$ 上的均匀分布. 模拟结果报告在表 4.3.1 和图 4.3.1 中, 表 4.3.1 提供了参数 $(\beta_1,\beta_2)^{\mathrm{T}}$ 估计的平均偏差 (Bias) 和标准差 (SD), 并比较了经验似然方法 (EL) 和基于渐近正态分布逼近的方法 (NA) 构造参数 $(\beta_1,\beta_2)^{\mathrm{T}}$ 的置信水平为 95%置信区间方面的表现, 表中提供了置信区间的置信下限 (CL)、置信上限 (CU)、平均长度 (AL) 和覆盖概率 (CP).

表 4.3.1 例 4.3.1 的模拟结果

n	β	估计效果		EL				NA			
		Bias	SD	CL	CU	AL	CP	CL	CU	AL	CP
$n=200$	β_1	0.0127	0.0833	0.9498	1.0480	0.0982	93.2%	0.9483	1.0553	0.1070	93.8%
	β_2	−0.0124	0.0535	−0.5167	−0.4840	0.0327		−0.5172	−0.4816	0.0356	
$n=400$	β_1	0.0052	0.0578	0.9660	1.0337	0.0677	94.4%	0.9625	1.0382	0.0757	94.4%
	β_2	−0.0044	0.0368	−0.5113	−0.4888	0.0225		−0.5125	−0.4873	0.0252	
$n=600$	β_1	0.0055	0.0447	0.9712	1.0292	0.0580	95.2%	0.9698	1.0312	0.0614	94.8%
	β_2	−0.0046	0.0282	−0.5095	−0.4903	0.0192		−0.5100	−0.4896	0.0204	

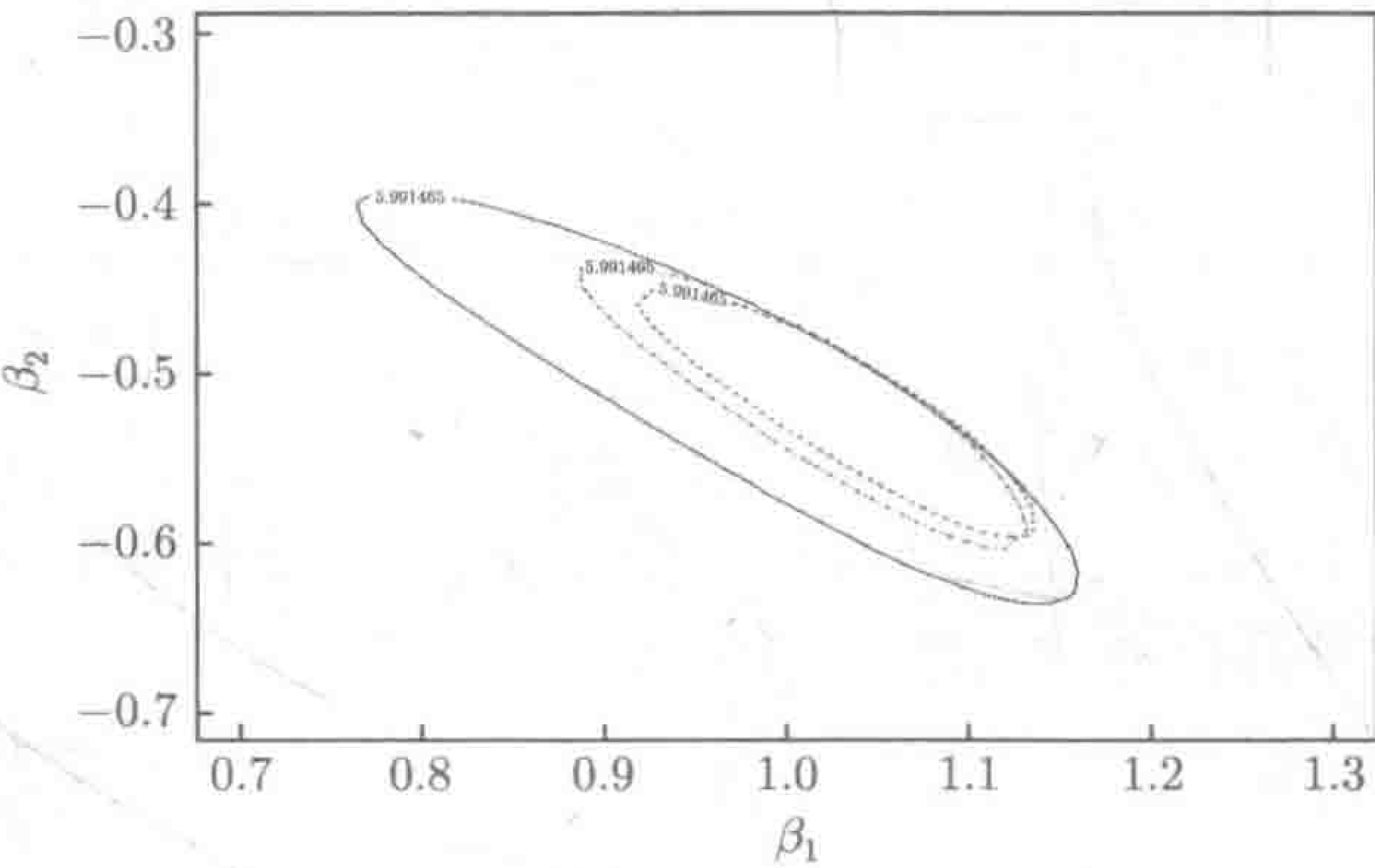


图 4.3.1 模型 (4.3.10) 中参数 $(\beta_1,\beta_2)^{\mathrm{T}}$ 的经验似然置信域, 实线表示 $n=200$ 时的置信域, 点虚线表示 $n=400$ 时的置信域, 虚线表示 $n=600$ 时的置信域, 图中 5.991465 表示 χ_2^2 分布的 95%下侧分位数

表 4.3.1 和图 4.3.1 的模拟结果说明以下三个方面:

(1) 从表 4.3.1 中的平均偏差和标准差可以看出, 随着样本量的增大, $(\beta_1,\beta_2)^{\mathrm{T}}$ 估计值越来越接近于真值 $(1,-0.5)^{\mathrm{T}}$;

(2) 对比于渐近正态分布逼近的方法, 经验似然方法构造的置信区间的平均区间长度更短, 覆盖概率更靠近真实的置信水平 95%;

(3) 图 4.3.1 展示了三种样本量, 分别为 200, 400 和 600 时, 经验似然方法构造的参数 $(\beta_1,\beta_2)^{\mathrm{T}}$ 置信水平为 95%的置信域. 从图中可以看出, 随着样本量的增大, $(\beta_1,\beta_2)^{\mathrm{T}}$ 的置信域变小.

例 4.3.2 本例取样本量 $n = 300$, 重复模拟 500 次. 数据集来自于下面非线性模型:

$$Y = \beta_1 + X_1(1 + X_2)^{\beta_2} + C_0 X_1 X_2 + \varepsilon, \tag{4.3.11}$$

这里的 C_0 选取为 0, 0.05, 0.10, 0.15, 0.20, 0.25, 0.30 和 0.35. 不难看出, 在 $C_0 = 0$ 时, 模型 (4.3.11) 变成了原假设 H_0 下的非线性模型. 本例中参数 β_1 和 β_2 的参数值设定和 X_1, X_2, ε 的随机数产生机制, 以及失真函数的选取都与例 4.3.1 完全相同, 模拟结果报告在表 4.3.2 中, 表中 α 表示检验水平.

表 4.3.2 例 4.3.2 的检验统计量功效值的比较

C_0	$\alpha = 0.01$	$\alpha = 0.025$	$\alpha = 0.05$	$\alpha = 0.10$
$C_0 = 0$	0.014	0.022	0.054	0.088
$C_0 = 0.05$	0.204	0.308	0.432	0.552
$C_0 = 0.10$	0.604	0.736	0.824	0.904
$C_0 = 0.15$	0.848	0.912	0.948	0.978
$C_0 = 0.20$	0.948	0.988	0.996	1.000
$C_0 = 0.25$	0.976	0.992	0.998	1.000
$C_0 = 0.30$	0.990	0.994	1.000	1.000
$C_0 = 0.35$	0.992	0.998	1.000	1.000

从表 4.3.2 中可以看出, 在 $C_0 = 0$ 时, 基于 wild-bootstrap 重抽样检验方法得到的 p 值非常接近于预先选定的 4 个检验水平 0.01, 0.025, 0.05, 0.10. 当 C_0 的取值逐渐变大时, wild-bootstrap 重抽样检验方法得到的 p 值越来越接近于 1. 表明 wild-bootstrap 重抽样检验方法是有效的.

4.3.6 实际数据分析

本节分析一组体重低于 1600 克的婴儿数据. 该数据集来自于美国杜克大学医学中心 Michael O'Shea 博士. 响应变量 Y 是婴儿体重 (BW), 协变量 X 是妊娠时间 (GA). 混淆变量 $(U_1, U_2)^{\mathrm{T}}$ 分别是婴儿出生前 4 天最低 pH(Lowest.pH) 和血小板数量 (Platelet. count). 图 4.3.2 展示了可加失真函数 $\phi_1(u_1)$ 和 $\phi_2(u_2)$ 估计的拟合曲线和 95%的置信带. 图 4.3.3 展示了可加失真函数 $\psi_1(u_1)$ 和 $\psi_2(u_2)$ 估计的拟合曲线和 95%的置信带. 这两个图都呈现了失真函数 $\phi(\cdot)$ 和 $\psi(\cdot)$ 都是非线性函数.

Zhang 等 (2014b) 把婴儿体重 (BW) 和妊娠时间 (GA) 之间的关系建立如下的非线性模型:

$$\mathrm{BW} = \beta_1 + (\beta_2 + \beta_3 \mathrm{GA})^{\beta_4} + \varepsilon. \tag{4.3.12}$$

使用 4.3 节介绍的估计方法, 可得参数向量 $(\beta_1, \beta_2, \beta_3, \beta_4)^{\mathrm{T}}$ 的估计为 $(\hat{\beta}_1, \hat{\beta}_2, \hat{\beta}_3, \hat{\beta}_4)^{\mathrm{T}} = (-0.228, -11.349, 0.485, 0.305)^{\mathrm{T}}$. 利用 4.3.2 节介绍的经验似然方法, 构造模型

中参数向量 $(\beta_1, \beta_2, \beta_3, \beta_4)^{\mathrm{T}}$ 中每一个参数分量的置信水平为 95% 的置信区间, 分别为

$$(-0.241, -0.214), \quad (-11.427, -11.267), \quad (0.482, 0.488), \quad (0.295, 0.316).$$

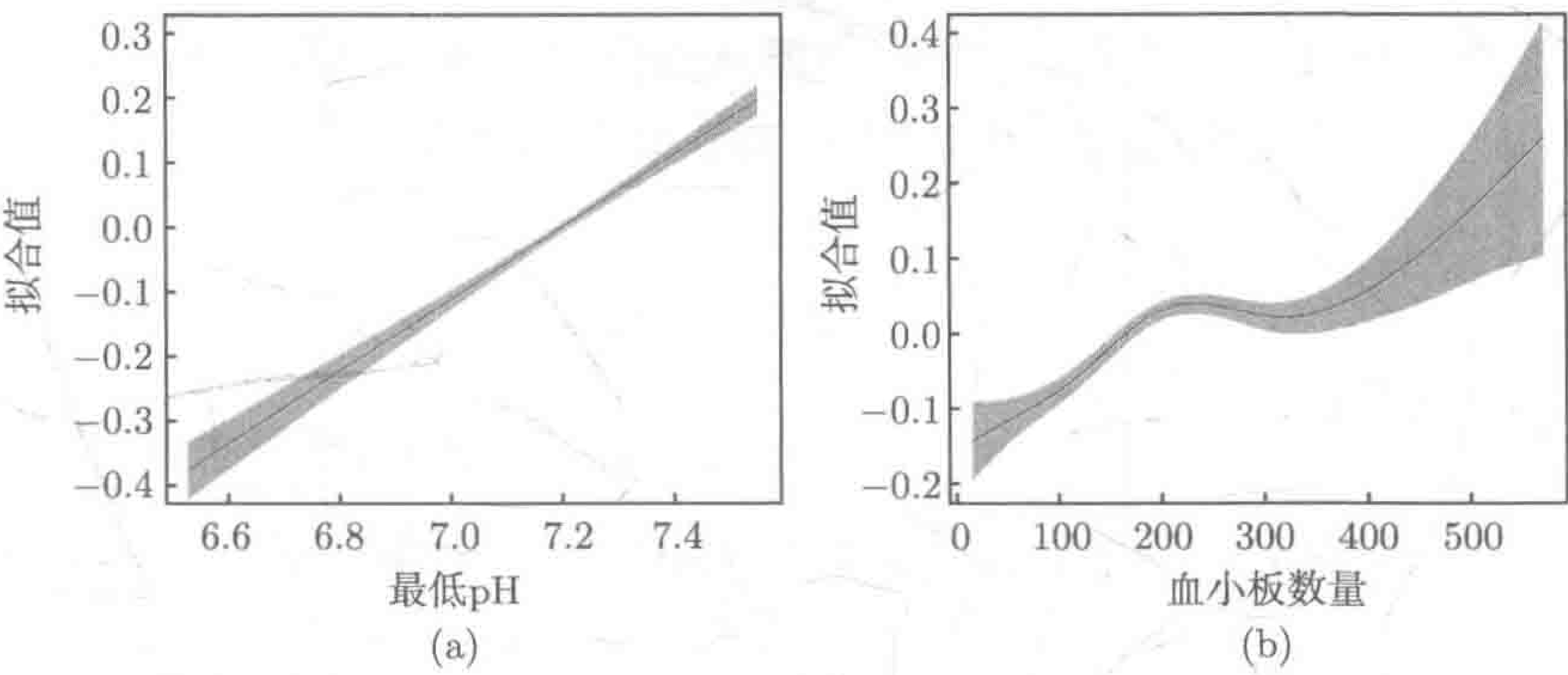


图 4.3.2 失真函数 $\phi_1(u_1)$ 和 $\phi_2(u_2)$ 估计的拟合曲线和 95% 的置信带, 实线表示估计的拟合曲线, 阴影部分表示 95% 的置信带. (a) 图表示 $\phi_1(u_1)$ 的拟合曲线和 95% 置信带, (b) 图表示 $\phi_2(u_2)$ 的拟合曲线和 95% 置信带

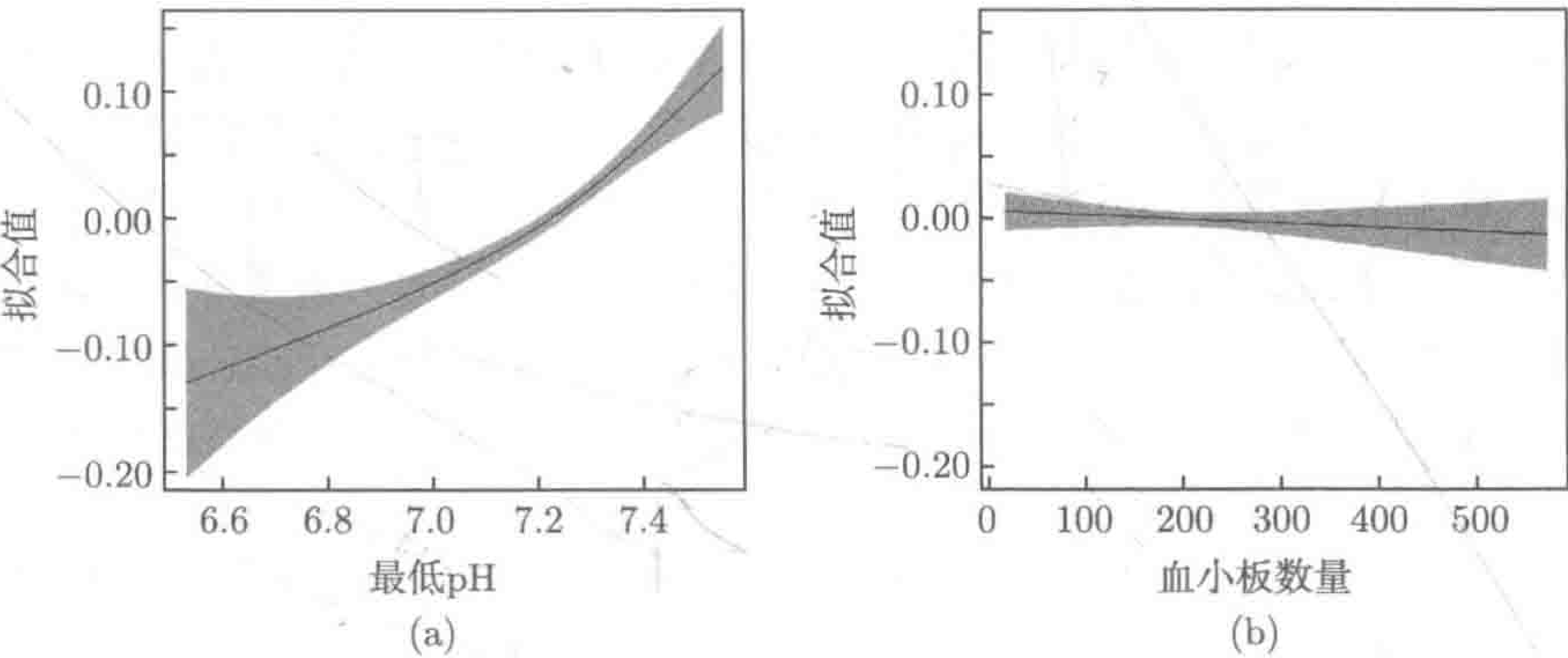


图 4.3.3 失真函数 $\psi_1(u_1)$ 和 $\psi_2(u_2)$ 估计的拟合曲线和 95% 的置信带, 实线表示估计的拟合曲线, 阴影部分表示 95% 的置信带. (a) 图表示 $\psi_1(u_1)$ 的拟合曲线和 95% 置信带, (b) 图表示 $\psi_2(u_2)$ 的拟合曲线和 95% 置信带

用 4.3.3 节提出的 Cramér-von Mises (CvM) 检验统计量来检验模型 (4.3.12) 是否可以很好地拟合这组数据, 即进行模型的拟合优度检验. 取 $N = 1000$, 计算 wild-bootstrap 检验统计量 $\mathcal{W}_{nb}$ 的 99%, 97.5%, 95%, 90% 的样本分位数分别为 1.7782, 0.0250, 0.0100 和 0.0077, 而计算得到的检验统计量 $\mathcal{W}_n$ 的值为 0.0076, 该数值结果表明模型 (4.3.12) 可以很好地拟合这组数据. 进一步, 图 4.3.4 提供了非

线性模型 (4.3.12) 的拟合曲线和置信水平为 95% 的置信带.

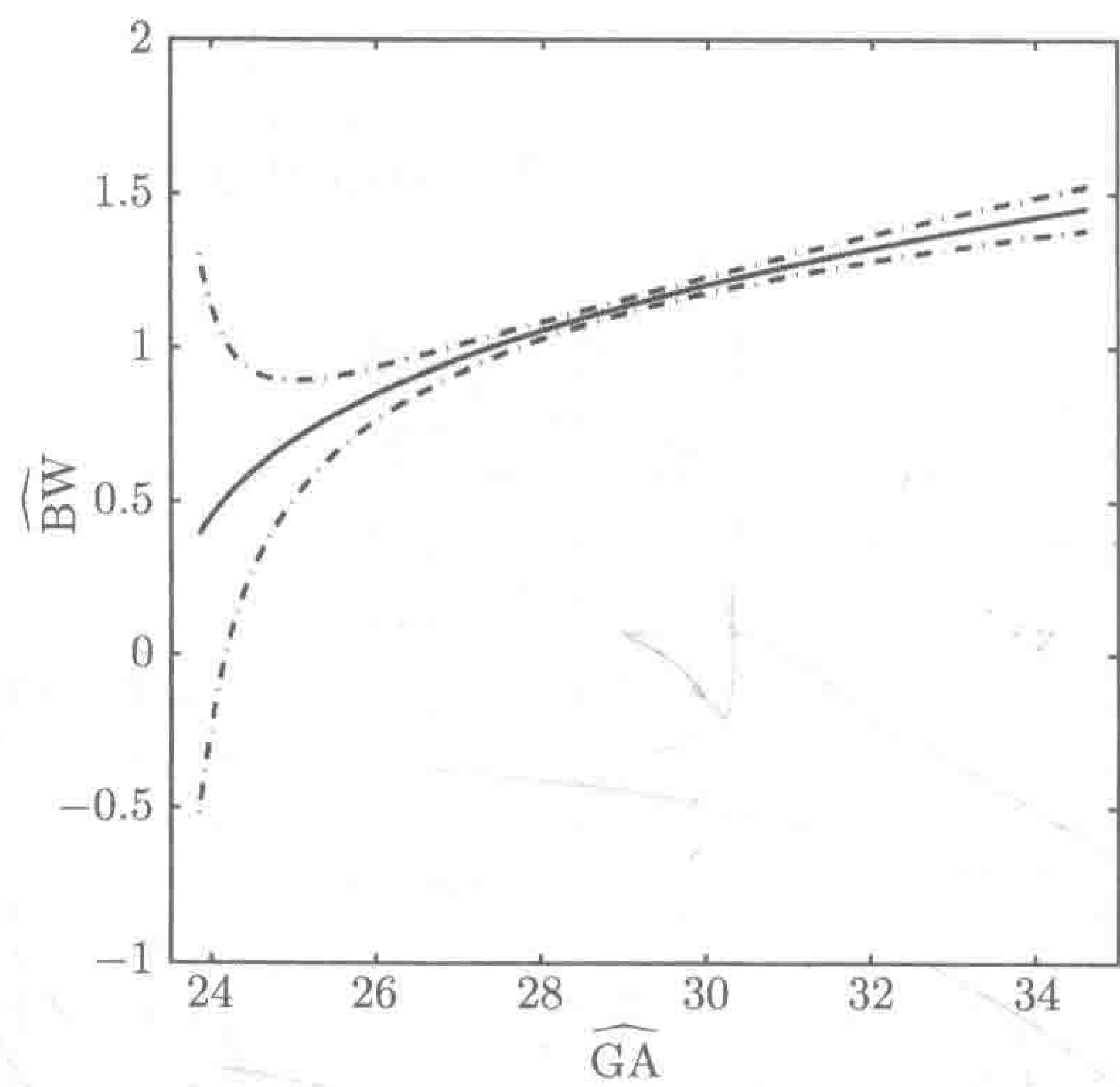


图 4.3.4　实线表示 $\widehat{\mathrm{BW}}$ 和 $\widehat{\mathrm{GA}}$ 之间的非线性模型 (4.3.12) 的拟合曲线, 点虚线表示置信水平为 95% 的置信带

第 5 章　部分线性测量误差模型

本章考虑两种测量误差的部分线性模型: 部分线性 EV 模型和协变量调整部分线性模型, 介绍这两种模型的估计方法、变量选择, 以及渐近结果.

5.1 引　言

5.1.1 部分线性模型介绍及研究现状

部分线性模型假设响应变量 Y 依赖于 p 维协变量 x 和一维协变量 T, 且 Y 与 x 之间呈线性关系, Y 与 T 之间呈非线性关系, 其模型形式为:

$$Y_i = x_i^{\mathrm{T}}\beta + g(T_i) + \varepsilon_i, \quad i = 1, \cdots, n, \tag{5.1.1}$$

其中 $\{x_i = (x_{i1}, \cdots, x_{ip})^{\mathrm{T}}, T_i, 1 \leqslant i \leqslant n\}$ 可以是随机设计也可以是固定设计. $\beta = (\beta_1, \cdots, \beta_p)^{\mathrm{T}}$ 是未知参数向量, $g(\cdot)$ 是一元未知函数, ε_i 是期望为 0 的随机误差.

自 Engle 等 (1986) 在研究气象条件对电力需求影响这一实际问题时首次提出部分线性模型以来, 该模型已出现了一系列丰富的研究成果. 由于部分线性模型结合了线性模型和非参数模型, 因此该模型在实际问题中有很强的解释能力, 并能够很好地拟合实际数据. 下面就部分线性模型的一些研究结果进行简单介绍.

Robinson (1988) 在非参数分量 $g(\cdot)$ 取 Nadaraga-Waston 核估计时, 构造了参数分量 β 的加权最小二乘估计 $\hat{\beta}$ 和非参数分量 $g(\cdot)$ 的估计 $\hat{g}(\cdot)$, 估计方法如下: 假设 β 已知, 定义非参数分量 $g(\cdot)$ 取 Nadaraga-Waston 核估计为

$$\hat{g}(t;\beta) = \sum_{i=1}^{n} W_{ni}(t)(Y_i - x_i^{\mathrm{T}}\beta), \tag{5.1.2}$$

其中 $W_{ni}(t) = \dfrac{K_h(T_i - t)}{\sum_{i=1}^{n} K_h(T_i - t)}$, $K_h(\cdot) = K(\cdot/h)/h$, 且 $K(\cdot)$ 表示核函数, h 表示一串收敛于 0 的常数列, 称为窗宽. 参数 β 的估计, 可以通过极小化下面的最小二乘目标函数得到:

$$Q(\beta) = \frac{1}{n}\sum_{i=1}^{n}[Y_i - x_i^{\mathrm{T}}\beta - \hat{g}(T_i;\beta)]^2. \tag{5.1.3}$$

简单计算, 可得参数 β 如下定义的最小二乘估计

$$\hat{\beta} = \left\{ \sum_{i=1}^{n} [x_i - \hat{g}_x(T_i)][x_i - \hat{g}_x(T_i)]^{\mathrm{T}} \right\}^{-1} \sum_{i=1}^{n} [x_i - \hat{g}_x(T_i)][Y_i - \hat{g}_Y(T_i)], \quad (5.1.4)$$

其中

$$\hat{g}_x(t) = \sum_{i=1}^{n} W_{ni}(t) x_i, \quad \hat{g}_Y(t) = \sum_{i=1}^{n} W_{ni}(t) Y_i$$

分别是 $g_x(t) = E(x|T = t)$ 和 $g_Y(t) = E(Y|T = t)$ 的 Nadaraga-Waston 核估计. 由式 (5.1.2) 和式 (5.1.4), 则有非参数 $g(\cdot)$ 的估计为

$$\hat{g}(t) = \sum_{i=1}^{n} W_{ni}(t)(Y_i - x_i^{\mathrm{T}} \hat{\beta}). \quad (5.1.5)$$

在一些正则条件下, Robinson (1988) 研究了估计 $\hat{\beta}$ 的渐近正态性及 $\hat{\beta}$ 和 $\hat{g}(\cdot)$ 的弱收敛速度.

Speckman (1988) 采用了参数化形式 Wr 逼近非参数分量 $g(\cdot)$, 其中 W 为某个给定的 $n \times q$ 的满秩矩阵, r 是附加的 $q \times 1$ 的未知参数向量, 部分线性模型 (5.1.1) 可以用矩阵形式表示为

$$Y = \mathbb{X}\beta + Wr + \varepsilon,$$

其中 $Y = (Y_1, \cdots, Y_n)^{\mathrm{T}}$, $\mathbb{X} = (x_1, \cdots, x_n)^{\mathrm{T}}$ 为 $n \times p$ 的矩阵, $\varepsilon = (\varepsilon_1, \cdots, \varepsilon_n)^{\mathrm{T}}$. 考虑同时极小化

$$\|Y - \mathbb{X}\beta - Wr\|^2 = \min, \quad \beta \in \mathbb{R}^p, \quad r \in \mathbb{R}^q,$$

可得 β 的估计

$$\hat{\beta} = (\mathbb{X}^{\mathrm{T}}(I - P_W)\mathbb{X})^{-1}\mathbb{X}^{\mathrm{T}}(I - P_W)Y, \quad (5.1.6)$$

其中 $P_W = W(W^{\mathrm{T}}W)^{-1}W^{\mathrm{T}}$ 为投影阵. 在适当的条件下, Speckman (1988) 研究了该估计的渐近性质. 由于在 β 估计的构造上已经消除了 T 对 x 和 Y 的影响, 因此 β 的估计是渐近无偏的.

Green 等 (1985) 提出可以用任一光滑矩阵 S (不必为投影阵) 替代式 (5.1.6) 中的 P_W, 由此可得 β 的估计

$$\widetilde{\beta} = (\mathbb{X}^{\mathrm{T}}(I - S)\mathbb{X})^{-1}\mathbb{X}^{\mathrm{T}}(I - S)Y.$$

由于这种估计是由 Green, Jennison 和 Seheult 提出的, 因此也称之为 G.J.S. 估计.

Engle 等 (1986) 使用光滑样条方法定义了 β 和 $g(\cdot)$ 的惩罚估计量为极小化

$$\frac{1}{n} \sum_{i=1}^{n} \{Y_i - x_i^{\mathrm{T}}\beta - g(T_i)\}^2 + \lambda \int \{g''(u)\}^2 \mathrm{d}u$$

的解, 其中 λ 是一个惩罚参数, 它起到在拟合程度与光滑程度之间的平衡作用. 这种方法既考虑到估计量同数据的拟合, 又顾及非参数分量估计的光滑性. Heckman (1986) 在 x_i 和 T_i 独立的情况下研究了 β 的惩罚最小二乘估计的相合性和渐近正态性. Rice (1986), Green 和 Silverman (1994), Eubank 等 (1998) 也分别在不同的条件下对估计的大样本性质进行了研究.

Hamilton 和 Truong (1997) 采用局部线性回归方法构造了参数和非参数分量的估计, 并证明了估计量的渐近正态性. Mammen 和 van de Geer (1997) 应用经验过程理论构造了惩罚拟似然估计, 并给出了该估计的渐近性质.

Ma 等 (2006) 研究了异方差的部分线性回归模型, 通过对加权估计方程进行修正, 即

$$0=\frac{1}{n}\sum_{i=1}^{n}\omega(x_i,T_i)\{Y_i-x_i^{\mathrm{T}}\beta-\hat{g}(T_i;\beta)\}\left[x_i-\frac{\hat{E}\{\omega(x,T)X|T_i\}}{\hat{E}\{\omega(x,T)|T_i\}}\right], \tag{5.1.7}$$

其中 $\omega(x_i,T_i)$ 表示权函数, 通常使用逆方差 (inverse-to-variation) 权, 即 $\omega(x_i,T_i)=\mathrm{Var}(\varepsilon_i|x_i,T_i)^{-1}$, $\hat{g}(T_i;\beta)$ 表示 $g(\cdot)$ 的非参数估计, $\hat{E}\{\omega(x,T)X|T_i\}$ 和 $\hat{E}\{\omega(x,T)|T_i\}$ 分别表示 $E\{\omega(x,T)X|T_i\}$ 和 $E\{\omega(x,T)|T_i\}$ 的非参数相合估计. 研究发现: 当 $\hat{g}$ 被错误指定时, 式 (5.1.7) 的右边仍可收敛到 0, 甚至错误地取 $\hat{g}\equiv 0$, 解方程 (5.1.7) 仍可得到相合估计 $\hat{\beta}$, 该估计具有渐近正态性, 并且是半参数有效估计.

Severini 和 Staniswalis (1994) 与 Härdle 等 (1998) 研究了模型 (5.1.1) 的推广形式, 即广义部分线性模型:

$$E(Y_i|x_i,T_i)=H\{x_i^{\mathrm{T}}\beta+g(T_i)\},\quad i=1,\cdots,n,$$

其中 H 为一个已知函数, 称为联系函数. 为了估计 β 和 $g(\cdot)$, Severini 和 Staniswalis (1994) 引进了拟似然估计方法, 该方法有类似于似然函数的性质, 但仅需指定 Y 的二阶矩而不是完全分布. 基于拟似然估计方法, Härdle 等 (1998) 考虑了 $g(\cdot)$ 的线性检验问题, 他们对检验问题的研究补充了 Severini 和 Staniswalis (1994) 的工作.

我国统计学者在部分线性模型的研究方面也作了大量相当深刻的工作. Liang (1992) 系统地研究了多种场合下 β 的渐近有效估计的构造. Shi (1992) 利用分块多项式逼近方法得到了 β 和 $g(\cdot)$ 的稳健 M 估计 $\hat{\beta}$ 和 $\hat{g}(\cdot)$, 在一定条件下证明了 $\hat{\beta}$ 具有渐近正态性, 并得到了 $\hat{\beta}$ 和 $\hat{g}(\cdot)$ 的弱收敛速度. 柴根象和徐克军 (1997), 钱伟民和柴根象 (1999), 钱伟民等 (2000) 应用小波方法研究部分线性模型, 建立了回归系数 β 和未知函数 $g(\cdot)$ 的小波估计, 证明了它们具有优良的大样本性质, 讨论了误差方差的小波估计及其渐近性质.

以上研究大都是在 (x_i^{T},T_i) 是随机设计点列的情况下进行讨论的, 当 (x_i^{T},T_i) 是固定设计时的研究成果相对较少. 而非随机设计情形并不是随机设计情形的特

例, 随机设计情况下的结果往往也不能简单地推广到非随机设计情形, 因为二者的处理方法和假设条件也有区别. 胡舒合 (1994), Gao 和 Zhao (1993), 高集体等 (1994, 1995) 分别研究了当 $g(\cdot)$ 的估计取一类非参数权函数估计时, β 的最小二乘估计和加权最小二乘估计的强相合性、渐近正态性、收敛速度, Berry-Esseen 界限以及重对数律等方面的大样本性质. 王启华 (1995), 王启华和郑忠国 (1997) 在截断样本下研究了 β 和 $g(\cdot)$ 的估计的强相合性, $p(p \geqslant 2)$ 阶平均相合性和渐近正态性. 陈明华 (1998a, 1998b) 讨论了 β 和 $g(\cdot)$ 的估计的强相合性, $p(p \geqslant 2)$ 阶平均相合性和收敛速度.

柴根象和孙平 (1995) 基于部分线性模型的可加性, 提出了新的二阶段估计方法, 得到了 β 和 $g(\cdot)$ 的核权函数形式的估计量 $\hat{\beta}$ 和 $\hat{g}(\cdot)$. 在 x_i 为固定设计点列而 T_i 为随机设计点列的情形下, 证明了 $\hat{g}(\cdot)$ 的强相合性和一致强相合性, 其一致强收敛速度可达到非参数回归函数估计的最优一致强收敛速度 $(n^{-1}\log n)^{\frac{1}{3}}$, 同时得到了 $\hat{\beta}$ 的渐近正态性.

薛留根 (2002, 2003) 将随机加权方法应用于部分线性模型的研究, 证明了用随机加权统计量的分布逼近原估计量误差的分布的强有效性, 并给出了估计量的最优收敛速度和随机加权逼近速度. 薛留根 (2004, 2005a) 研究了部分线性模型中误差方差估计之分布的非一致收敛速度和 Berry-Esseen 界.

近些年来, 对部分线性模型的研究一直是统计学界的一大热点, 其研究不断向各方面发展. 例如, 对删失数据、缺失数据、纵向数据、面板数据、测量误差数据、高维或超高维数据、异方差、相依混合序列等情形下的研究都取得了可喜的研究成果. 关于部分线性模型方法的详细讨论可进一步参考专著柴根象和洪圣岩 (1995), Härdle 等 (2000), 薛留根 (2012), 李高荣和杨宜平 (2015).

5.1.2　部分线性可加测量误差模型

在许多实际问题中, 对于模型 (5.1.1) 中的响应变量 $\{Y_i\}$ 和协变量 $\{x_i, T_i\}$, 经常不能直接观测或准确测量, 相应测量到的是带有测量误差的响应变量和协变量. 通常把模型 (5.1.1) 中变量带有测量误差的模型称为部分线性测量误差模型或部分线性 EV(errors-in-variables) 模型. 对响应变量 Y_i 具有测量误差时比较简单, 可以把测量误差吸收到模型误差中进行处理. 大部分文献集中在协变量带有测量误差的情形, 即考虑如下的部分线性 EV 模型:

$$\begin{cases} Y = x^{\mathrm{T}}\beta + g(T) + \varepsilon, \\ X = x + u, \\ V = T + v, \end{cases} \tag{5.1.8}$$

其中 (X, V) 为取值于 $\mathbb{R}^p \times [0,1]$ 上的可观测随机向量, x 为 $\mathbb{R}^p$ 上不可观测的随机

向量或称为潜在变量 (latent variable), T 为 [0,1] 上不可观测的随机变量或潜在变量, β 为 $p\times 1$ 未知参数向量, $g(\cdot)$ 是定义于闭区间 $[0,1]$ 上的未知函数, 模型误差 ε 与 (X,V) 相互独立, 且均值为 0, 方差为 σ^2. 测量误差 u 是独立同分布的向量, 与 (Y,x,T) 独立, 均值为 0, 协方差阵为 Σ_{uu}. 测量误差 v 是独立同分布的随机变量, 与 (Y,x,T) 独立, 均值为 0, 方差为 σ_v^2. 考虑到模型的可识别性, 假设测量误差的协方差阵 Σ_{uu} 已知, 测量误差 v 有一个已知的分布. 模型 (5.1.8) 它表明变量 Y 关于 (x,T) 的回归函数 $E(Y|(x,T))$ 呈偏线性的形式, 且变量 x 和 T 不能直接观测到, 所能观测到的是分别受了误差变量 u 和 v 干扰后的变量 X 和 V.

在模型 (5.1.8) 中, 如果变量 T 被准确观测, 则表示只有线性部分的协变量 x 具有测量误差, 这时 $\sigma_v^2=0$. 如果协变量 x 被准确观测, 则表示只有非参数部分的协变量 T 具有测量误差, 这时协方差矩阵 Σ_{uu} 为零矩阵, 即 $\Sigma_{uu}=0$. 许多统计学者对模型 (5.1.8) 进行了深入研究, 对该模型的统计推断问题主要集中在对 β 和 $g(\cdot)$ 的估计方面. 例如, 崔恒建 (1995b) 在测量误差向量为椭球等高分布且方差阵已知的情形下获得了诸多参数和非参数估计及其性质; Liang 等 (1999) 在测量误差向量协方差阵已知, 且设计向量为随机设计的情形下, 获得了诸多参数和非参数估计及其渐近性质; 李高荣等 (2006) 利用经验似然方法获得了部分线性 EV 模型中参数的极大经验似然估计和非参数 $g(\cdot)$ 估计量的最优弱收敛速度, 并且证明了极大经验似然估计的渐近正态性; Li 和 Xue (2008) 研究了部分线性测量误差模型的经验似然统计推断问题, 证明了经验对数似然比统计量的渐近分布并构造了模型中兴趣参数分量的置信域; Liang (2000), Zhu 和 Cui (2003) 考虑了非参数部分协变量也带有测量误差的情形; Wang (1999) 与 Wang 和 Zhu (2001) 则考虑了部分线性 EV 模型中具有删失数据和带有核实数据情形时的参数和非参数估计, 并获得了它们的渐近性质; 崔恒建 (2004) 考虑有重复观测的部分线性 EV 模型的参数估计问题.

5.2 校正的估计方法

5.2.1 参数部分带有测量误差情况

本节集中考虑协变量 T 准确测量, 即 $\sigma_v^2=0$, 而协变量 x 具有可加测量误差的部分线性 EV 模型. 假设 $\{X_i=(X_{i1},\cdots,X_{ip})^{\mathrm T},T_i,Y_i,1\leqslant i\leqslant n\}$ 为来自模型 (5.1.8) 的一组独立同分布的可观测随机样本, 即考虑如下的部分线性 EV 模型:

$$\begin{cases} Y_i=x_i^{\mathrm T}\beta+g(T_i)+\varepsilon_i, \\ X_i=x_i+u_i, \quad i=1,\cdots,n. \end{cases} \tag{5.2.1}$$

Liang 等 (1999) 在假定协方差矩阵 Σ_{uu} 已知情况下, 讨论了模型 (5.2.1) 中参

数和非参数的估计问题. Liang 等 (1999) 证明, 在式 (5.1.4) 定义的参数 β 的估计中, 如果忽略测量误差直接用观测变量 X_i 代替 x_i, 则导致所得的参数 β 的估计是不相合的. 为了克服由测量误差引起参数估计的不相合性, Liang 等 (1999) 提出了校正衰减 (correction attenuation) 的修正思想, 主要想法如下. 注意到测量误差 u 独立于变量 (Y, x, T), 且 $E(u) = 0$ 和 $\mathrm{Cov}(u) = \Sigma_{uu}$, 如果直接用观测变量 X 代替 x, 则总体表示的最小二乘目标函数为

$$E[Y - X^{\mathrm{T}}\beta - g(T)]^2 = E[Y - x^{\mathrm{T}}\beta - g(T)]^2 + \beta^{\mathrm{T}}\Sigma_{uu}\beta. \tag{5.2.2}$$

因此, 为了获得参数 β 的相合估计, Liang 等 (1999) 构造了下面校正的最小二乘目标函数

$$Q(\beta) = \frac{1}{n}\sum_{i=1}^{n}\{Y_i - X_i^{\mathrm{T}}\beta - \hat{g}(T_i;\beta)\}^2 - \beta^{\mathrm{T}}\Sigma_{uu}\beta,$$

其中 $\hat{g}(T_i;\beta)$ 是假定 β 已知时 $g(\cdot)$ 的核估计, 定义见式 (5.1.2). 在校正的最小二乘目标函数 $Q(\beta)$ 中减去一项 $\beta^{\mathrm{T}}\Sigma_{uu}\beta$, 主要目的是衰减由测量误差带来的影响, 保证校正的最小二乘目标函数在真值附近衰减到 0, 使得求解校正的最小二乘目标函数得到的解是相合的. 极小化校正的最小二乘目标函数 $Q(\beta)$, 得到 β 的相合估计量为

$$\begin{aligned}\hat{\beta}_n = &\left[\sum_{i=1}^{n}\{X_i - \hat{g}_X(T_i)\}\{X_i - \hat{g}_X(T_i)\}^{\mathrm{T}} - n\Sigma_{uu}\right]^{-1}\\ &\times\sum_{i=1}^{n}\{X_i - \hat{g}_X(T_i)\}\{Y_i - \hat{g}_Y(T_i)\},\end{aligned} \tag{5.2.3}$$

其中 $\hat{g}_X(\cdot)$ 和 $\hat{g}_Y(\cdot)$ 分别是 $E(X|T)$ 和 $E(Y|T)$ 的核估计, 定义为

$$\hat{g}_X(t) = \sum_{i=1}^{n} W_{ni}(t)X_i, \quad \hat{g}_Y(t) = \sum_{i=1}^{n} W_{ni}(t)Y_i.$$

Liang 等 (1999) 要求权函数 $W_{ni}(t)$ 为一般的权函数, 需要满足:

$$\max_{1\leqslant i\leqslant n}\sum_{j=1}^{n} W_{nj}(T_i) = O_P(1), \quad \max_{1\leqslant i,j\leqslant n}\sum_{j=1}^{n} W_{nj}(T_i) = O_P(b_n),$$

$$\max_{1\leqslant i\leqslant n}\sum_{j=1}^{n} W_{nj}(T_i)I(|T_j - T_i| > c_n) = O_P(c_n),$$

其中 $b_n = n^{-4/5}, c_n = n^{-1/5}\log n$. 通常取权函数 $W_{ni}(t)$ 为 Nadaraga-Waston 权函数和 Gasser 和 Müller (1979) 定义的如下 Gasser-Müller 权函数

$$W_{ni}(t) = \frac{1}{h}\int_{s_{i-1}}^{s_i} K\left(\frac{t-s}{h}\right)\mathrm{d}s, \quad i = 1, \cdots, n,$$

其中 $s_0=0, s_n=1, s_i=(T_{(i)}+T_{(i+1)})/2, 1\leqslant i\leqslant n-1, T_{(i)}$ 为 T_i 的次序统计量, $1\leqslant i\leqslant n$. $K(\cdot)$ 为 $[-1,1]$ 上有界且对称的概率密度函数, h 为收敛于 0 的窗宽. 注意到

$$g(t)=E[(Y-x^{\mathrm{T}}\beta)|T=t]=E[(Y-X^{\mathrm{T}}\beta)|T=t].$$

由上式, 式 (5.2.3) 和式 (5.1.2), 则得到 $g(t)$ 的最终估计为

$$\hat{g}_n(t)=\sum_{i=1}^{n}W_{ni}(t)(Y_i-X_i^{\mathrm{T}}\hat{\beta}_n).$$

由式 (5.2.2), 则有 $\sigma^2=E[Y_i-X_i^{\mathrm{T}}\beta-g(T_i)]^2-\beta^{\mathrm{T}}\Sigma_{uu}\beta$. 由参数和非参数的估计量, 可定义模型误差方差 σ^2 的估计为

$$\hat{\sigma}_n^2=\frac{1}{n}\sum_{i=1}^{n}[Y_i-X_i^{\mathrm{T}}\hat{\beta}_n-\hat{g}_n(T_i)]^2-\hat{\beta}_n^{\mathrm{T}}\Sigma_{uu}\hat{\beta}_n.$$

Liang 等 (1999) 给出了下面三个定理研究了参数估计 $\hat{\beta}_n$, 方差估计 $\hat{\sigma}_n^2$ 和非参数估计 $\hat{g}_n(t)$ 的渐近性质. 定理的详细证明, 请读者参考 Liang 等 (1999).

定理 5.2.1 假设 $\sup\limits_{0\leqslant t\leqslant 1}E(\|x\|^4|T=t)<\infty$, $E(\varepsilon^4+\|u\|^4)<\infty$, $B=E(x-E(x|T))^{\otimes 2}$ 为正定矩阵, 当 $n\to\infty$ 时, 则有

$$\sqrt{n}(\hat{\beta}_n-\beta)\xrightarrow{L}N(0,B^{-1}\Gamma B^{-1}),$$

其中 $\Gamma=E[(\varepsilon-u^{\mathrm{T}}\beta)\{x-E(x|T)\}]^{\otimes 2}+E\{(uu^{\mathrm{T}}-\Sigma_{uu})\beta\}^{\otimes 2}+E(uu^{\mathrm{T}}\varepsilon^2)$. 如果 ε 为同方差且独立于 (x,T), 则 $\Gamma=E[(\varepsilon-u^{\mathrm{T}}\beta)]^2B+E\{(uu^{\mathrm{T}}-\Sigma_{uu})\beta\}^{\otimes 2}+\sigma^2\Sigma_{uu}$.

定理 5.2.2 假设定理 5.2.1 的条件成立, 并假设 ε_i 为同方差且独立于 (x_i,T_i), 且有 $\mathrm{Var}(\varepsilon_i)=\sigma^2$, 当 $n\to\infty$ 时, 则有

$$\sqrt{n}(\hat{\sigma}_n^2-\sigma^2)\xrightarrow{L}N(0,\sigma_*^2),$$

其中 $\sigma_*^2=E\{(\varepsilon-u^{\mathrm{T}}\beta)^2-(\beta^{\mathrm{T}}\Sigma_{uu}\beta+\sigma^2)\}^2$.

为了对 β 进行统计推断, 需要估计 $\hat{\beta}_n$ 的渐近方差, B 和 Γ 的相合估计分别如下:

$$\hat{B}=\frac{1}{n-p}\sum_{i=1}^{n}[X_i-\hat{g}_X(T_i)]^{\otimes 2}-\Sigma_{uu}, \tag{5.2.4}$$

$$\hat{\Gamma}=\frac{1}{n}\sum_{i=1}^{n}\left\{[X_i-\hat{g}_X(T_i)][Y_i-X_i^{\mathrm{T}}\hat{\beta}_n-\hat{g}_n(T_i)]+\Sigma_{uu}\hat{\beta}_n\right\}^{\otimes 2}. \tag{5.2.5}$$

在 ε_i 同方差情况下, 即有方差 σ^2, 并模型误差 ε_i 独立于 (x_i,T_i,u_i), u 为正态分布, 则 Γ 的相合估计定义如下:

$$(\hat{\sigma}_n^2+\hat{\beta}_n^{\mathrm{T}}\Sigma_{uu}\hat{\beta}_n)\hat{B}+E[(uu^{\mathrm{T}}-\Sigma_{uu})\hat{\beta}_n]^{\otimes 2}+\hat{\sigma}_n^2\Sigma_{uu}.$$

在许多实际问题中, Σ_{uu} 已知的假定是不成立的, 因此需要估计 Σ_{uu} 的大小. Liang 等 (1999) 建议采用重复测量的方法来估计测量误差的方差阵 Σ_{uu}. 假定重复测量, 即 $X_{ij}=x_i+u_{ij}, j=1,\cdots,m_i$. 令 $\bar{X}_i=\dfrac{1}{m_i}\sum\limits_{j=1}^{m_i}X_{ij}$ 表示重复测量的样本均值, 则 Σ_{uu} 的一个无偏的矩估计可以定义为

$$\hat{\Sigma}_{uu}=\frac{\sum\limits_{i=1}^{n}\sum\limits_{j=1}^{m_i}(X_{ij}-\bar{X}_i)(X_{ij}-\bar{X}_i)^{\mathrm{T}}}{\sum\limits_{i=1}^{n}(m_i-1)}. \tag{5.2.6}$$

关于测量误差协方差阵 Σ_{uu} 的估计问题, 详细的讨论见 Liang 等 (1999).

定理 5.2.3 假设定理 5.2.1 的条件成立, 又假设 $W_{ni}(t)$ 满足 1 阶 Lipschitz 条件, $i=1,\cdots,n$, 当 $n\to\infty$ 时, 则对固定的 t, 估计量 $\hat{g}_n(t)$ 的渐近偏差和渐近方差分别为 $\sum\limits_{i=1}^{n}W_{ni}(t)g(T_i)-g(t)$ 和 $\sum\limits_{i=1}^{n}W_{ni}^2(t)(\beta^{\mathrm{T}}\Sigma_{uu}\beta+\sigma^2)$. 特别地, 对核估计, 二者的阶都为 $O(n^{-2/5})$.

5.2.2 非参数部分带有测量误差情况

对于模型 (5.1.8), Liang (2000) 考虑了协变量 x 被准确测量, 即 $\Sigma_{uu}=0$, 而协变量 T 具有可加测量误差的部分线性 EV 模型. 假设 $\{x_i=(x_{i1},\cdots,x_{ip})^{\mathrm{T}},V_i,Y_i,1\leqslant i\leqslant n\}$ 为来自模型 (5.1.8) 的一组独立同分布的可观测随机样本, 即考虑如下的部分线性 EV 模型:

$$\begin{cases} Y_i=x_i^{\mathrm{T}}\beta+g(T_i)+\varepsilon_i, \\ V_i=T_i+v_i, \quad i=1,\cdots,n, \end{cases} \tag{5.2.7}$$

其中测量误差 v_i 是 i.i.d. 的, 并独立于 (Y_i,x_i,T_i), 均值为 0, 方差为 σ_v^2. 为了保证模型的可识别性, 假设测量误差 v_i 有一个已知的分布函数.

当 T_i 能够被准确观测时, Robinson (1988) 已经证明式 (5.1.4) 定义的参数分量 β 的估计是相合的, 并具有渐近正态性. 但是如果 T_i 具有测量误差, 这时式 (5.1.4) 定义的参数分量 β 的估计不再是相合的和渐近有效的. 原因很简单, 对于式 (5.1.4) 中的 $\hat{g}_x(T_i)$ 和 $\hat{g}_Y(T_i)$, 如果直接用观测变量 V_i 代替不可观测的变量 T_i, 肯定会导致这两个函数的估计不再是精确和有效的. 此外, 受到非参数部分测量误差的影响, 由式 (5.1.5) 定义的非参数函数 $g(\cdot)$ 的估计也不是渐近有效的. 因此, 需要发展新的校正方法对参数分量和非参数分量进行有效估计.

假定 β 已知, 模型 (5.2.7) 变为非参数 EV 模型, 这时可借助 Fan 和 Truong (1993) 提出的逆卷积的方法首先对非参数函数 $g(\cdot)$ 进行校正估计. 令 $f_V(\cdot)$ 和 $f_T(\cdot)$

分别表示 V 和 T 的密度函数, $f_T(\cdot)$ 的估计可以定义为

$$\hat{f}_T(t) = \frac{1}{nh}\sum_{i=1}^{n} K_n\left(\frac{V_i - t}{h}\right), \tag{5.2.8}$$

其中

$$K_n(t) = \frac{1}{2\pi}\int_{\mathbb{R}} \exp(-\mathrm{i}st)\frac{\phi_K(s)}{\phi_v(s/h)}\mathrm{d}s, \tag{5.2.9}$$

这里 $\phi_K(\cdot)$ 表示核函数 $K(\cdot)$ 的 Fourier 变换, $\phi_v(\cdot)$ 是误差变量 v 的特征函数. 令

$$\omega_{ni}(\cdot) = K_n\left(\frac{\cdot - V_i}{h}\right)\Big/\sum_{j=1}^{n} K_n\left(\frac{\cdot - V_j}{h}\right) =: \frac{1}{nh}K_n\left(\frac{\cdot - V_i}{h}\right)\Big/\hat{f}_T(\cdot). \tag{5.2.10}$$

由 $g(t) = E(Y - x^{\mathrm{T}}\beta|T = t)$, Liang (2000) 定义了 $g(\cdot)$ 如下的"伪估计":

$$g_n(t;\beta) = \sum_{i=1}^{n} \omega_{ni}(t)(Y_i - x_i^{\mathrm{T}}\beta). \tag{5.2.11}$$

在最小二乘目标函数 (5.1.3) 中的 $\hat{g}(T_i;\beta)$ 用"伪估计" $g_n(V_i;\beta)$ 替换, 然后关于 β 求解最小二乘目标函数的极小值, 得到 β 的最小二乘估计为

$$\hat{\beta}_n = \left\{\sum_{i=1}^{n} \widetilde{x}_i\widetilde{x}_i^{\mathrm{T}}\right\}^{-1}\sum_{i=1}^{n} \widetilde{x}_i\widetilde{Y}_i. \tag{5.2.12}$$

其中

$$\widetilde{x}_i = x_i - \sum_{j=1}^{n} \omega_{nj}(V_i)x_j, \quad \widetilde{Y}_i = Y_i - \sum_{j=1}^{n} \omega_{nj}(V_i)Y_j.$$

最后, 由式 (5.2.11) 和式 (5.2.12), 可得 $g(\cdot)$ 的最终估计如下

$$\hat{g}_n(t) = \sum_{i=1}^{n} \omega_{ni}(t)(Y_i - x_i^{\mathrm{T}}\hat{\beta}_n). \tag{5.2.13}$$

为了得到估计量的渐近结果, 需要下面的假设条件.

(C5.2.1) 假设 $\sup\limits_{0\leqslant t\leqslant 1} E(\|x_1\|^3|T = t) < \infty$, 且 $B = E(x - E(x|T))^{\otimes 2}$ 为正定矩阵.

(C5.2.2) 假设 $g(\cdot)$, $g_x(t)$ 和 $g_Y(t)$ 是 1 阶 Lipschitz 连续的.

(C5.2.3) (1) 密度函数 $f_T(\cdot)$ 在区间 [0,1] 上是有界远离 0 的, 且有有界的 k 阶导数, 其中 k 是一个整数.

(2) 测量误差 v 分布的特征函数 $\phi_v(\cdot)$ 是非退化的.

(3) 测量误差 v 的分布函数是超光滑的或是普通光滑的. 下面给出 Fan 和 Truong (1993) 定义的超光滑和普通光滑的定义.

① α 阶超光滑: 如果测量误差分布的特征函数 $\phi_v(\cdot)$ 满足

$$d_0|t|^{\alpha_0}\exp(-|t|^{\alpha}/\zeta)\leqslant|\phi_v(t)|\leqslant d_1|t|^{\alpha_1}\exp(-|t|^{\alpha}/\zeta),\quad \text{当 } t\to\infty\text{时},\tag{5.2.14}$$

其中 d_0,d_1,α 和 ζ 是正的常数, α_0 和 α_1 是常数.

② α 阶普通光滑: 如果测量误差分布的特征函数 $\phi_v(\cdot)$ 满足

$$d_0|t|^{-\alpha}\leqslant|\phi_v(t)|\leqslant d_1|t|^{-\alpha},\quad \text{当 } t\to\infty\text{时},\tag{5.2.15}$$

其中 d_0,d_1,α 为正常数.

(C5.2.4) 核函数 $K(\cdot)$ 是 k 阶核函数, 即

$$\int_{-\infty}^{\infty}K(u)\mathrm{d}u=1,\quad \int_{-\infty}^{\infty}u^lK(u)\mathrm{d}u\begin{cases}=0, & l=1,\cdots,k-1,\\ \neq 0, & l=k.\end{cases}$$

定理 5.2.4 假设条件 (C5.2.1) ~条件 (C5.2.4) 成立, $E(|\varepsilon|^3+\|u\|^3)<\infty$. 如果下面两条中任何一个成立, 即

(1) 测量误差分布为 α 阶超光滑. $\phi_K(t)$ 在 $|t|\leqslant M_0$ 上有一个有界的支撑, 且取窗宽 $h=c(\log n)^{-1/\alpha}$, 其中 $c>M_0(2/\zeta)^{1/\alpha}$;

(2) 测量误差分布为 α 阶普通光滑. 对一些常数 $c\neq 0$, 有

$$t^{\alpha}\phi_v(t)\to c,\quad t^{\alpha+1}\phi_v'(t)=O(1),\quad t\to\infty$$

和

$$\int_{-\infty}^{\infty}|t|^{\alpha+1}\{\phi_K(t)+\phi_K'(t)\}\mathrm{d}t<\infty,\quad \int_{-\infty}^{\infty}|t^{\alpha+1}\phi_K(t)|^2\mathrm{d}t<\infty.$$

取窗宽 $h=dn^{-1/(2k+2\alpha+1)}, d>0$.

当 $n\to\infty$ 时, $\hat{\beta}_n$ 是渐近正态的, 即

$$\sqrt{n}(\hat{\beta}_n-\beta)\xrightarrow{L}N(0,\sigma^2B^{-1}).$$

定理 5.2.4 的详细证明, 请读者参考 Liang (2000).

5.2.3 参数和非参数部分都带有测量误差情况

本节考虑参数和非参数部分都带有测量误差的部分线性 EV 模型 (5.1.8), 主要介绍 Zhu 和 Cui (2003) 提出的矩估计和逆卷积的校正估计方法, 构造了参数和非参数的估计.

为了构造估计量, 首先介绍一些概念. 令

$$\begin{aligned}H(X,T) &= X - E(X|T) = x - E(x|T) + u,\\ H(Y,T) &= Y - E(Y|T) = [x - E(x|T)]^{\mathrm{T}}\beta + \varepsilon,\end{aligned}$$

并令 $\varpi(t) \geqslant 0$ 是在支撑 [0,1] 上的权函数. 为了避免核估计分母所引起的边界问题, 对 T 的密度函数有如下的假设:

$$0 < \inf_{0\leqslant t\leqslant 1} f_T(t) \leqslant \sup_{0\leqslant t\leqslant 1} f_T(t) < \infty.$$

进一步, 令

$$\begin{aligned}S_1 &= E[H(X,T)H^{\mathrm{T}}(X,T)\varpi(T)]\\ &= E\Big\{[x - E(x|T)][x - E(x|T)]^{\mathrm{T}}\varpi(T)\Big\} + E(\varpi(T))\Sigma_{uu}\\ &=: S + S_3\Sigma_{uu}, \qquad (5.2.16)\\ S_2 &= E[H(X,T)H(Y,T)\varpi(T)] = S\beta,\\ S_4 &= E[(\varepsilon - u^{\mathrm{T}}\beta)^2\varpi(T)],\end{aligned}$$

其中

$$S = E\Big\{[x - E(x|T)][x - E(x|T)]^{\mathrm{T}}\varpi(T)\Big\}, \quad S_3 = E(\varpi(T)).$$

令 $f(y,x,t)$ 表示 (Y,X,T) 的密度函数, 且令

$$\begin{aligned}g_X(t) &= E(X|T=t) =: (g_{X,1}(t),\cdots,g_{X,p}(t))^{\mathrm{T}},\\ g_Y(t) &= E(Y|T=t).\end{aligned} \qquad (5.2.17)$$

如果 S 是一个正定矩阵, 由式 (5.2.16) 和式 (5.2.17), 可得 $\beta, g(t)$ 和 σ^2 的表达式如下:

$$\begin{aligned}\beta &= (S_1 - S_3\Sigma_{uu})^{-1}S_2,\\ g(t) &= g_Y(t) - g_X(t)^{\mathrm{T}}\beta,\\ \sigma^2 &= \frac{S_4}{S_3} - \beta^{\mathrm{T}}\Sigma_{uu}\beta.\end{aligned} \qquad (5.2.18)$$

这时 $\beta, g(t)$ 和 σ^2 的估计问题转化为 $S_1, S_2, S_3, S_4, g_X(t)$ 和 $g_Y(t)$ 的估计问题, Zhu 和 Cui (2003) 提出了下面校正的矩估计方法. 假设样本 $\{(X_i, Y_i, V_i), i = 1,\cdots,n\}$ 是 i.i.d. 的随机样本, 来自于模型 (5.1.8), 估计程序如下.

步骤 1 由于非参数部分协变量也具有测量误差, 则由逆卷积的核估计思想, 可用式 (5.2.8) 定义密度函数 $f_T(\cdot)$ 的逆卷积核估计, 记为 $\hat{f}_T(t)$.

步骤 2　定义 (Y,X,T) 的联合密度函数 $f(y,x,t)$, $g_X(t)$ 和 $g_Y(t)$ 的核估计分别如下:

$$\hat{f}_n(y,x,t)=\frac{1}{nh^{p+2}}\sum_{i=1}^{n}\prod_{k=1}^{p}K\left(\frac{x_k-X_{ik}}{h}\right)K\left(\frac{y-Y_i}{h}\right)K_n\left(\frac{t-V_i}{h}\right),$$
$$\hat{g}_X(t)=\sum_{i=1}^{n}\omega_{ni}(t)X_i,\quad \hat{g}_Y(t)=\sum_{i=1}^{n}\omega_{ni}(t)Y_i,$$

其中 $(y,x,t)\in\mathbb{R}\times\mathbb{R}^p\times\mathbb{R}$, $K_n(\cdot)$ 由式 (5.2.9) 所定义, 权函数 $\omega_{ni}(\cdot)$ 由式 (5.2.10) 所定义, $K(\cdot)$ 是一个核函数, h 是窗宽. Stefanski 和 Carroll (1990), Fan 和 Truong (1993) 在一定的正则条件下, 证明了 $\hat{f}_n(y,x,t),\hat{g}_X(t)$ 和 $\hat{g}_Y(t)$ 分别是 $f(y,x,t)$, $g_X(t)$ 和 $g_Y(t)$ 的相合估计.

步骤 3　由式 (5.2.16), 构造 $S_i, i=1,2,3$, β 和 $g(\cdot)$ 的估计如下

$$\begin{cases}\hat{S}_1=\displaystyle\int_{\mathbb{R}^p}\int_{\mathbb{R}}\int_{\mathbb{R}}(x-\hat{g}_X(t))\,(x-\hat{g}_X(t))^{\mathrm{T}}\,\varpi(t)\hat{f}_n(y,x,t)\mathrm{d}x\mathrm{d}y\mathrm{d}t,\\ \hat{S}_2=\displaystyle\int_{\mathbb{R}^p}\int_{\mathbb{R}}\int_{\mathbb{R}}(x-\hat{g}_X(t))\,(y-\hat{g}_Y(t))\,\varpi(t)\hat{f}_n(y,x,t)\mathrm{d}x\mathrm{d}y\mathrm{d}t,\\ \hat{S}_3=\displaystyle\int_{\mathbb{R}}\varpi(t)\hat{f}_T(t)\mathrm{d}t,\end{cases}$$

$$\hat{\beta}=(\hat{S}_1-\hat{S}_3\Sigma_{uu})^{-1}\hat{S}_2,\quad \hat{g}(t)=\hat{g}_Y(t)-\hat{g}_X(t)^{\mathrm{T}}\hat{\beta}.$$

步骤 4　构造 S_4 和 σ^2 的估计如下

$$\hat{S}_4=\int_{\mathbb{R}^p}\int_{\mathbb{R}}\int_{\mathbb{R}}\left(y-x^{\mathrm{T}}\hat{\beta}-\hat{g}(t)\right)^2\varpi(t)\hat{f}_n(y,x,t)\mathrm{d}x\mathrm{d}y\mathrm{d}t,$$
$$\hat{\sigma}^2=\frac{\hat{S}_4}{\hat{S}_3}-\hat{\beta}^{\mathrm{T}}\Sigma_{uu}\hat{\beta}.$$

Zhu 和 Cui (2003) 在测量误差分布为 α 阶普通光滑和一些正则条件下, 证明了参数 β 和 σ^2 估计 $\hat{\beta}$ 和 $\hat{\sigma}^2$ 的相合性和渐近正态性, 以及非参数函数 $g(\cdot)$ 估计 $\hat{g}(\cdot)$ 的相合性和收敛速度. 对于定理的详细讨论和证明, 读者可参考 Zhu 和 Cui (2003).

5.3　经验似然方法

5.3.1　引言

为了获得参数 β 估计的精度, 还需要构造参数 β 的置信域. 对于这个问题通常的方法是: ① 首先用 β_n 去估计 β; ② 构造 β_n 的渐近方差的相合估计; ③ 利用渐

近正态分布来构造未知参数的置信域. 然而有时在半参数和非参数情形, 方差的估计是相当困难的, 尤其对带有测量误差的半参数和非参数情形; 此外, 利用渐近正态分布得到的置信域是对称的, 而实际数据的分布可能是不对称的, 所以这是不恰当的. 而利用经验似然方法构造 β 的置信域可以避免对渐近方差进行估计, 而且所得到的置信域的形状由数据自行决定. 因此一些统计学家把这一方法应用于各种统计模型的研究. 例如, Owen (1991) 将经验似然方法应用到线性模型, 并且证明了经验对数似然比渐近于一个 χ^2 变量, 因此可以直接应用它来构造回归系数的大样本置信区间或置信域; Kolaczyk (1994) 与 Chen 和 Cui (2003) 将经验似然方法进一步推广到了广义线性模型之中; Wang 和 Jing (1999), Shi 和 Lau (2000) 把经验似然方法应用于部分线性模型的研究; Li 和 Xue (2008) 把经验似然方法用于参数部分带有测量误差的部分线性 EV 模型的研究, 提出了校正的经验似然方法, 构造了模型中参数分量的经验似然置信域; Wong 等 (2009) 研究了异方差的部分线性 EV 模型的经验似然诊断问题; Yan 和 Chen (2014) 把校正的经验似然方法推广到参数和非参数部分都带有测量误差的部分线性 EV 模型的研究; 更多的研究成果可见参考文献.

本节把经验似然方法用于部分线性 EV 模型 (5.2.1) 的研究, 利用纠偏思想通过对辅助随机向量进行修正, 提出了一种校正的经验似然方法, 使得所构造的模型中未知参数的经验对数似然比函数依分布收敛到 χ^2 分布, 所得结果能够构造参数 β 的置信域. 进一步证明了所得到的极大经验似然估计与最小二乘估计有相同的渐近方差. 由模拟可知在置信域的构造方面经验似然方法优于最小二乘方法. 此外还给出了非参数估计量的最优收敛速度. 此外当假定模型误差 ε_i 是同方差时, 获得了共同方差的估计, 并证明了该估计的渐近正态性.

5.3.2 经验似然方法和渐近性质

假设 $\{X_i=(X_{i1},\cdots,X_{ip})^{\mathrm{T}},T_i,Y_i,1\leqslant i\leqslant n\}$ 为来自模型 (5.2.1) 的一组独立同分布的可观测随机样本. 首先假设 β 是已知的, 则模型 (5.2.1) 可看成一个非参数回归模型:

$$Y_i-x_i^{\mathrm{T}}\beta=g(T_i)+\varepsilon_i,\quad i=1,\cdots,n. \tag{5.3.1}$$

由 $g(t)=E(Y_i-x_i^{\mathrm{T}}\beta|T_i=t)=E(Y_i-X_i^{\mathrm{T}}\beta|T_i=t)$, 所以 $g(T_i)$ 可以由 $Y_i-X_i^{\mathrm{T}}\beta$ 和 T_i 使用下列方法进行估计: 如核估计法、局部多项式估计法、近邻估计法、光滑样条估计方法、B 样条估计方法和小波估计等非参数方法. Fan (1993), Fan 和 Gijbels (1996) 提出的局部多项式光滑方法具有良好的性质. 如它可以减少 Nadaraya-Watson 估计的偏差和 Gasser-Müller 估计的方差; 此外局部多项式光滑方法还能适应边界效应. 所以本节使用局部线性光滑方法来构造 $g(\cdot)$ 的估计量.

如果 T_i 在 t 的一个小邻域内, 可以用一个线性函数局部地逼近回归函数 $g(T_i)$, 即

$$g(T_i) \approx g(t) + g'(t)(T_i - t) =: a + b(T_i - t).$$

首先假定 β 已知, 求 a 和 b, 可以通过极小化下面的加权局部最小二乘问题

$$\sum_{i=1}^{n}\{Y_i - X_i^{\mathrm{T}}\beta - a - b(T_i - t)\}^2 K_h(T_i - t), \tag{5.3.2}$$

其中 $K_h(\cdot) = K(\cdot/h)/h$, $K(\cdot)$ 是一个核函数, h 是一串收敛于 0 的常数列, 称之为窗宽. 记 $\hat{a}$ 和 $\hat{b}$ 是加权局部最小二乘问题 (5.3.2) 的解. 简单计算得

$$\hat{a} = \frac{\sum_{j=1}^{n} w_j(t)(Y_j - X_j^{\mathrm{T}}\beta)}{\sum_{j=1}^{n} w_j(t)}, \tag{5.3.3}$$

其中

$$w_j(t) = K_h(T_j - t)[S_{n,2}(t) - (T_j - t)S_{n,1}(t)],$$
$$S_{n,l}(t) = \frac{1}{n}\sum_{j=1}^{n} K_h(T_j - t)(T_j - t)^l, \quad l = 0, 1, 2.$$

用 $\hat{g}(t;\beta)$ 记 $g(t)$ 的“伪估计”, 即

$$\hat{g}(t;\beta) = \sum_{j=1}^{n} W_{nj}(t)(Y_j - X_j^{\mathrm{T}}\beta), \tag{5.3.4}$$

其中

$$W_{nj}(t) = w_j(t)\Big/\sum_{j=1}^{n} w_j(t).$$

令

$$\widetilde{Y}_i = Y_i - \sum_{j=1}^{n} W_{nj}(T_i)Y_j, \quad \widetilde{x}_i = x_i - \sum_{j=1}^{n} W_{nj}(T_i)x_j,$$
$$\widetilde{X}_i = X_i - \sum_{j=1}^{n} W_{nj}(T_i)X_j, \quad \widetilde{g}_i = g(T_i) - \sum_{j=1}^{n} W_{nj}(T_i)g(T_j).$$

现在引入一个辅助随机向量

$$\eta_i(\beta) = \widetilde{X}_i(Y_i - X_i^{\mathrm{T}}\beta - \hat{g}(T_i;\beta)) + \Sigma_{uu}\beta = \widetilde{X}_i(\widetilde{Y}_i - \widetilde{X}_i^{\mathrm{T}}\beta) + \Sigma_{uu}\beta. \tag{5.3.5}$$

注意到为了避免由测量误差所引起的对参数的过低估计, 在辅助变量 $\eta_i(\beta)$ 的构造上, 式 (5.3.5) 右边加了一项 $\Sigma_{uu}\beta$ 来进行修正. 当 β 是真实参数时, $E(\eta_i(\beta))=o(1)$. 反之, 当 $E(\eta_i(\beta))=o(1)$ 时, 可以构造一个估计方程 $\sum\limits_{i=1}^{n}\eta_i(\beta)=0$. 因此, 利用信息 $E(\eta_i(\beta))=o(1)$, 提出下面的经验对数似然比函数

$$l(\beta)=\max\left\{\sum_{i=1}^{n}\log(np_i)\;\middle|\;p_i\geqslant 0,\ \sum_{i=1}^{n}p_i=1,\ \sum_{i=1}^{n}p_i\eta_i(\beta)=0\right\}. \tag{5.3.6}$$

利用 Lagrange 乘子法, 令

$$G=\sum_{i=1}^{n}\log p_i+n\log n+t(\beta)\left(1-\sum_{i=1}^{n}p_i\right)-n\lambda^{\mathrm{T}}(\beta)\sum_{i=1}^{n}p_i\eta_i(\beta). \tag{5.3.7}$$

由 $\dfrac{\partial G}{\partial p_i}=0$, 以及 $\sum\limits_{i=1}^{n}p_i\dfrac{\partial G}{\partial p_i}=0$ 得

$$p_i=\frac{1}{n}\frac{1}{1+\lambda^{\mathrm{T}}(\beta)\eta_i(\beta)},\quad i=1,\cdots,n. \tag{5.3.8}$$

由式 (5.3.8), $l(\beta)$ 可以重新表示为

$$l(\beta)=-\sum_{i=1}^{n}\log(1+\lambda^{\mathrm{T}}(\beta)\eta_i(\beta)), \tag{5.3.9}$$

其中 $\lambda=\lambda(\beta)$ 为 Lagrange 乘子, 满足

$$\frac{1}{n}\sum_{i=1}^{n}\frac{\eta_i(\beta)}{1+\lambda^{\mathrm{T}}\eta_i(\beta)}=0. \tag{5.3.10}$$

极大化 $l(\beta)$ 可得 β 的极大经验似然估计 (MELE) $\widetilde{\beta}$. 需要注意: 如果 $l(\beta)$ 在 $\hat{\beta}$ 处极大化, 并且有 $\hat{p}_i(\hat{\beta})=\dfrac{1}{n}(1\leqslant i\leqslant n)$, 则 $\hat{\beta}=\left(\sum\limits_{i=1}^{n}\widetilde{X}_i\widetilde{X}_i^{\mathrm{T}}-n\Sigma_{uu}\right)^{-1}\sum\limits_{i=1}^{n}\widetilde{X}_i\widetilde{Y}_i$. 由 MELE $\widetilde{\beta}$, 以及式 (5.3.4), 可以定义 $g(\cdot)$ 的估计为

$$\widetilde{g}_n(t)=\sum_{j=1}^{n}W_{nj}(t)(Y_j-X_j^{\mathrm{T}}\widetilde{\beta}). \tag{5.3.11}$$

当假定模型误差 ε_i 是同方差时, 且共同的方差为 σ^2. 由于

$$E\{Y_i-x_i^{\mathrm{T}}\beta-g(T_i)\}^2=\sigma^2$$

和

$$E\{Y_i - X_i^{\mathrm{T}}\beta - g(T_i)\}^2 = E\{Y_i - x_i^{\mathrm{T}}\beta - g(T_i)\}^2 + \beta^{\mathrm{T}}\Sigma_{uu}\beta,$$

所以由 MELE $\widetilde{\beta}$ 和式 (5.3.11) 可得 σ^2 的估计为

$$\hat{\sigma}^2 = \frac{1}{n}\sum_{i=1}^{n}(\widetilde{Y}_i - \widetilde{X}_i^{\mathrm{T}}\widetilde{\beta})^2 - \widetilde{\beta}^{\mathrm{T}}\Sigma_{uu}\widetilde{\beta}. \tag{5.3.12}$$

记

$$h_j(T_i) = E(x_{ij}|T_i), \quad \xi_i = x_i - E(x_i|T_i), \quad 1 \leqslant i \leqslant n, \quad 1 \leqslant j \leqslant p,$$

其中 $x_i = (x_{i1}, \cdots, x_{ip})^{\mathrm{T}}$, 并且 x_{ij} 表示 x_i 的第 j 个分量. 在下面条件中用 $\|\cdot\|$ 表示 Euclidean 范数. 为了得到本节的渐近结果, 需要如下的正则条件:

(C5.3.1) T_1 在 $[0,1]$ 上有连续密度函数 $f(t)$, 且

$$0 < \inf_{0\leqslant t\leqslant 1} f(t) \leqslant \sup_{0\leqslant t\leqslant 1} f(t) < \infty.$$

(C5.3.2) $g(t)$ 在 $(0,1)$ 上具有有界的二阶连续偏导数; $h_j(\cdot)(1 \leqslant j \leqslant p)$ 满足 1 阶 Lipschitz 条件, 即存在常数 $M > 0$, 使对任意 $t_1, t_2 \in [0,1]$, $|h_j(t_1) - h_j(t_2)| \leqslant M|t_1 - t_2|$, 对 $1 \leqslant j \leqslant p$ 均成立.

(C5.3.3) 核函数 $K(\cdot)$ 是对称的, 非负的, 并在其支撑集 $[-1,1]$ 上是连续可微的, 满足 1 阶 Lipschitz 条件, 并且 $\int_{-1}^{1} K(y)\mathrm{d}y = 1$, $\int_{-1}^{1} y^2K(y)\mathrm{d}y \neq 0$.

(C5.3.4) 窗宽 h 满足: 当 $n \to \infty$ 时, $h \to 0$, $nh^2/\log^2 n \to \infty$, $nh^4 \to 0$.

(C5.3.5) $\sup\limits_{0\leqslant t\leqslant 1} E(\|x_1\|^4|T = t) < \infty$, $E\|\xi_i\|^4 < \infty$, $B = E(\xi_1\xi_1^{\mathrm{T}})$ 是正定阵.

(C5.3.6) $E(\varepsilon_i) = E(u_i) = 0$, $\sup\limits_i E(\varepsilon_i^4 + \|u_i\|^4) < \infty$.

定理 5.3.1　假设正则条件(C5.3.1)~条件(C5.3.6)成立, 当 $\beta = \beta_0$(真参数), $n \to \infty$ 时, 则

$$-2l(\beta_0) \xrightarrow{L} \chi_p^2.$$

基于定理 5.3.1, 对任给的 $0 < \alpha < 1$, 存在 c_α 使得 $P(\chi_p^2 > c_\alpha) = \alpha$, 则

$$I_\alpha(\beta^*) = \{\beta^* \in \mathbb{R}^p \mid -2l(\beta^*) \leqslant c_\alpha\}$$

为参数向量 β 的具有渐近置信水平为 $1-\alpha$ 的置信域.

定理 5.3.2　假设正则条件(C5.3.1)~条件(C5.3.6)成立, 当 $n \to \infty$ 时, 则 $\widetilde{\beta}$ 是一个渐近正态估计, 即

$$\sqrt{n}(\widetilde{\beta} - \beta) \xrightarrow{L} N(0, B^{-1}\Gamma B^{-1}),$$

其中 $\Gamma = E[(\varepsilon - u^{\mathrm{T}}\beta)\{x - E(x|T)\}]^{\otimes 2} + E\{(uu^{\mathrm{T}} - \Sigma_{uu})\beta\}^{\otimes 2} + E(uu^{\mathrm{T}}\varepsilon^2)$. 如果 ε 是同方差并独立于 (x, T), 则 $\Gamma = E(\varepsilon - u^{\mathrm{T}}\beta)^2 B + E\{(uu^{\mathrm{T}} - \Sigma_{uu})\beta\}^{\otimes 2} + \Sigma_{uu}\sigma^2$, 其中 B 在条件(C5.3.5)中给出.

对比定理 5.2.1, 可以发现, β 的极大经验似然估计和最小二乘估计是渐近等价的, 即有相同的渐近方差. 为了利用定理 5.3.2 对 β 进行统计推断, 需要对渐近协方差进行估计. 类似于定理 5.2.1 中渐近协方差的估计方法, 只需要把式 (5.2.4) 和式 (5.2.5) 中的参数和非参数的估计分别用本节定义的极大经验似然估计 $\widetilde{\beta}$ 和由式 (5.3.11) 定义的非参数估计 $\widetilde{g}_n(t)$ 替换, 则可以得到渐近协方差中 B 和 Γ 的相合估计.

定理 5.3.3 假设正则条件(C5.3.1)~条件(C5.3.6)成立, 并且模型误差 ε_i 独立于 (x_i, T_i) 是同方差的, 共同的方差为 σ^2, 当 $n \to \infty$ 时, 则由式 (5.3.12) 定义的估计有渐近正态性, 即

$$\sqrt{n}(\hat{\sigma}^2 - \sigma^2) \xrightarrow{L} N(0, \sigma_*^2),$$

其中 $\sigma_*^2 = E\{(\varepsilon - u^{\mathrm{T}}\beta)^2 - (\beta^{\mathrm{T}}\Sigma_{uu}\beta + \sigma^2)\}^2$.

对比定理 5.2.2, 可以发现, 当模型误差 ε 为同方差, 并独立于 (x, T) 时, σ^2 的两种估计量是渐近等价的, 即有相同的渐近方差.

定理 5.3.4 假设正则条件(C5.3.1)~条件(C5.3.6)成立, 当 $n \to \infty$ 时, 则 $\widetilde{g}_n(t)$ 是 $g(t)$ 的相合估计. 进一步地, 对固定的 $t \in [0,1]$, 当选取 $h = cn^{-1/3}$(c 为正的常数) 时, 有

$$\widetilde{g}_n(t) - g(t) = O_P(n^{-1/3}).$$

限于篇幅, 此处省去了定理 5.3.1~定理 5.3.4 的详细证明. 感兴趣的读者, 可参考李高荣等 (2006), 李高荣 (2007), Li 和 Xue (2008).

5.3.3 模拟研究

本节将进行模拟研究以说明所得的主要结果, 并和最小二乘方法进行比较. 还利用核估计和局部多项式估计对回归函数 $g(t)$ 在有限样本情况下的实际表现进行了比较.

考虑模型: $Y_i = 1 + \beta x_i + 4\sin(6T_i) + \varepsilon_i$, $X_i = x_i + u_i$, $i = 1, \cdots, n$, 这里 $\beta = 2$, $\varepsilon_i \sim N(0, 0.2^2)$, $x_i \sim N(0,1)$ 是受到测量误差 u_i 干扰的不可观测的一维随机向量, 其中 $u_i \sim N(0, \sigma_u^2)$, T_i 为服从区间 $[0,1]$ 上的均匀分布的随机变量. 在下面的模拟中, 取样本容量分别为 $n = 50, 100, 150, 200$, 测量误差方差分别取 $\sigma_u^2 = 0.2^2, 0.4^2, 0.6^2$, 考虑用经验似然方法 (EL) 和 Liang 等 (1999) 中的最小二乘方法 (LS), 通过模拟研究比较这两种方法在置信域的精度和覆盖概率大小方面的优劣, 以及极大经验似然估计 (MELE) 和最小二乘估计 (LSE).

核函数取为 Epanechnikov 核$K(u)=\frac{3}{4}(1-u^2)I(|u|\leqslant 1)$, 窗宽的选取采用欠光滑 (undersmoothing) 方法, 作法如下: 首先利用“去一分量”的交叉证实法选择最优窗宽 h_{opt}, 然后用 $n^{-2/15}$ 乘以 h_{opt}, 即得所选择的窗宽$\hat{h}_n=h_{\text{opt}}\cdot n^{-2/15}=O(n^{-1/3})$ 满足条件中对窗宽的要求. 重复模拟次数为 $M=1000$ 次, 关于 β 的两种估计, 分别计算

$$\bar{\beta}=\frac{1}{M}\sum_{i=1}^{M}\hat{\beta}_i,$$

$$\widehat{\text{Bias}}(\hat{\beta})=\bar{\beta}-\beta,$$

$$\widehat{\text{MSE}}(\hat{\beta})=\frac{1}{M-1}\sum_{i=1}^{M}(\hat{\beta}_i-\beta)^2,$$

进行比较, 说明两种估计方法的完成情况.

基于样本大小 $n=100$ 时函数 $g(t)=1+4\sin(6t)$的核估计曲线和局部多项式估计曲线如图 5.3.1 所示.

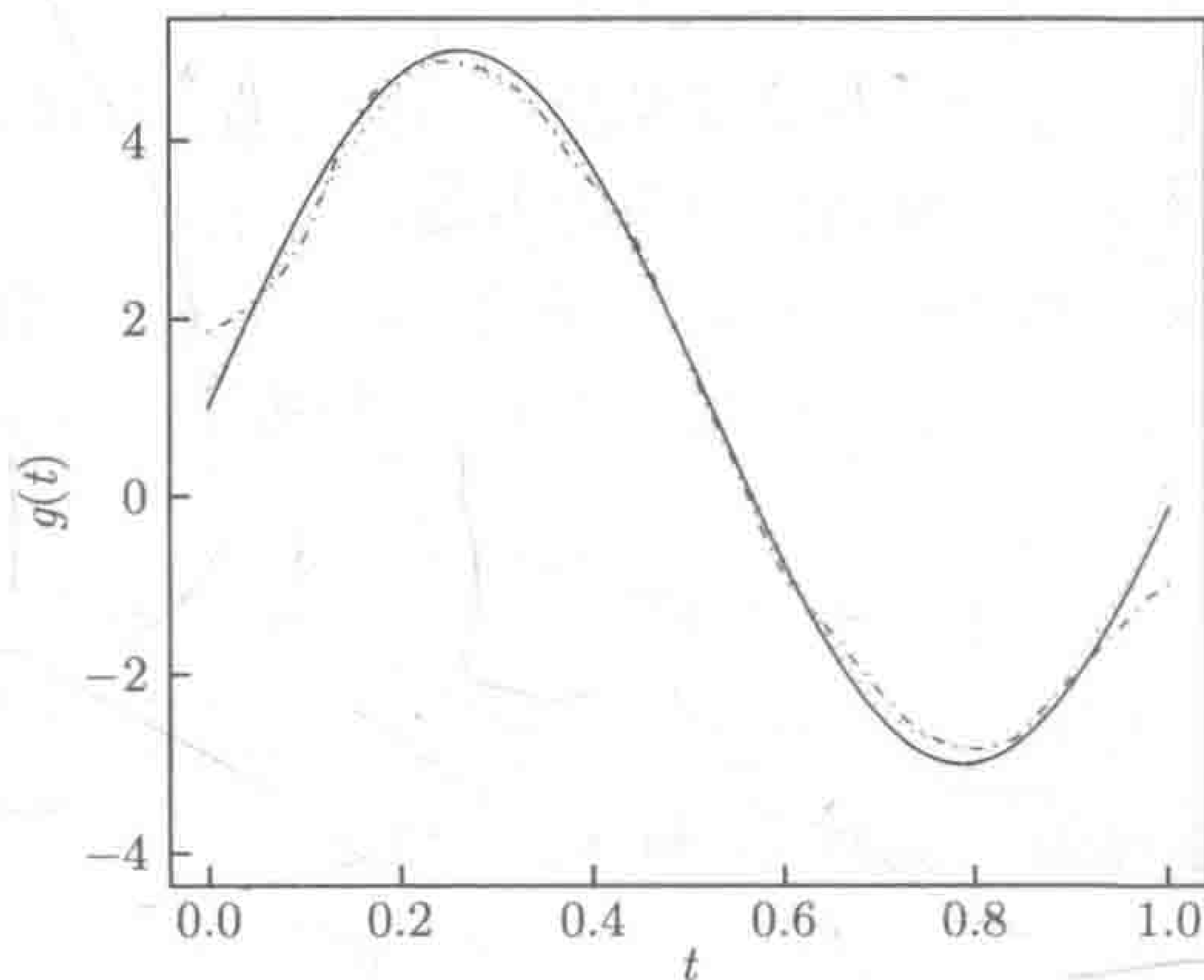


图 5.3.1　函数 $g(t)=1+4\sin(6t)$ 的真实曲线和两种估计曲线, 实线表示真实曲线, 虚线表示局部线性估计, 点断线表示核估计

从表 5.3.1 可以看出, 随着样本量的增加, EL 方法和 LS 方法所构造的置信区间的平均长度 (AL) 都不断减小, 而相应的覆盖概率则不断增加, 直至接近置信水平 0.95. 此外, 这两种方法都随着测量误差方差 σ_u^2 的增大, 而使得覆盖概率减小, 区间的平均长度增大. 还可以看到, EL 比 LS 方法的覆盖概率大, 平均区间长度短, 因此模拟结果说明 EL 比 LS 方法要好.

从表 5.3.2 可以看出, β 的 MELE 和 LSE 都有较好的拟合效果, 而且随着样本容量的增加, 对于不同的测量误差方差, β 的两种估计都接近于真值, 估计的偏差、均方差也相应减小. 并且从模拟来看, 这两种方法基本等价, 有相同的拟合效果. 此外也说明纠偏方法的有效性, 因为测量误差的方差对估计影响变化不是很大.

表 5.3.1 β 的置信水平为 95%的平均置信区间和覆盖概率

σ_u^2	n	覆盖概率		EL 的区间		LS 的区间	
		EL	LS	$(\beta_L,\beta_U)_{\mathrm{EL}}$	$\mathrm{AL}_{\mathrm{EL}}$	$(\beta_L,\beta_U)_{\mathrm{LS}}$	$\mathrm{AL}_{\mathrm{LS}}$
0.2^2	50	0.9310	0.9190	(1.8546,2.1097)	0.2551	(1.8733,2.1467)	0.2734
	100	0.9370	0.9280	(1.9203,2.0941)	0.1738	(1.9078,2.0922)	0.1844
	150	0.9460	0.9430	(1.9462,2.0957)	0.1495	(1.9447,2.0953)	0.1506
	200	0.9490	0.9470	(1.9398,2.0578)	0.1180	(1.9302,2.0601)	0.1299
0.4^2	50	0.9298	0.9160	(1.8501,2.1799)	0.3298	(1.8936,2.2839)	0.3903
	100	0.9360	0.9280	(1.9043,2.2042)	0.2999	(1.8791,2.2138)	0.3347
	150	0.9450	0.9410	(1.9129,2.1675)	0.2546	(1.8982,2.1827)	0.2845
	200	0.9520	0.9430	(1.8964,2.1266)	0.2302	(1.9156,2.1573)	0.2417
0.6^2	50	0.9260	0.9180	(1.8956,2.3143)	0.4187	(1.8734,2.3266)	0.4532
	100	0.9310	0.9330	(1.8216,2.2014)	0.3798	(1.8819,2.2772)	0.3953
	150	0.9430	0.9400	(1.8557,2.2105)	0.3548	(1.8262,2.1939)	0.3677
	200	0.9470	0.9450	(1.8799,2.1995)	0.3196	(1.9045,2.2303)	0.3258

表 5.3.2 在不同 σ_u^2 下 β 的 MELE 和 LSE 及其精度

σ_u^2	n	MELE			LSE		
		$\bar{\beta}$	$\widehat{\mathrm{Bias}}(\hat{\beta})$	$\widehat{\mathrm{MSE}}(\hat{\beta})$	$\bar{\beta}$	$\widehat{\mathrm{Bias}}(\hat{\beta})$	$\widehat{\mathrm{MSE}}(\hat{\beta})$
0.2^2	50	2.0024	0.0024	0.00179	1.9977	−0.0023	0.00239
	100	2.0012	0.0012	0.00151	1.9988	−0.0012	0.00147
	150	1.9991	−0.0009	0.00128	2.0007	0.0007	0.00137
	200	2.0002	0.0002	0.00097	2.0005	0.0005	0.00119
0.4^2	50	2.0028	0.0028	0.00399	1.9975	−0.0025	0.00373
	100	1.9982	−0.0018	0.00304	2.0017	0.0017	0.00324
	150	2.0014	0.0014	0.00248	2.0015	0.0015	0.00302
	200	1.9991	−0.0009	0.00235	1.9993	−0.0007	0.00242
0.6^2	50	1.9955	−0.0045	0.00531	2.0034	0.0034	0.00641
	100	2.0039	0.0039	0.00433	2.0023	0.0023	0.00429
	150	1.9978	−0.0022	0.00351	1.9983	−0.0017	0.00384
	200	2.0016	0.0016	0.00271	2.0012	0.0012	0.00276

图 5.3.1 表明函数 $g(t)$ 的两种估计 (即核估计和局部线性估计) 在内点的拟合效果都很好, 但在边界点 0 和 1 附近, 核估计表现要差, 而局部线性估计则很好地处理了边界效应.

5.4 两种变量选择方法

在数据分析中, 首先要判断哪些协变量对响应变量是显著的, 应该保留在最终的统计模型中, 同时要判断哪些变量不是显著的, 从模型中排除. 目前, 变量选择方

法是统计建模和数据分析的一个基本工具. Liang 和 Li (2009) 研究了模型 (5.2.1) 的变量选择问题, 提出了两种变量选择方法, 分别是校正的惩罚最小二乘方法和惩罚分位数回归方法. 本节主要介绍 Liang 和 Li (2009) 的这两种变量选择程序.

5.4.1 校正的惩罚最小二乘方法

类似于 5.3 节定义的概念, 令 $\widetilde{Y}_i = Y_i - \sum_{j=1}^{n} W_{nj}(T_i)Y_j$, $\widetilde{X}_i = X_i - \sum_{j=1}^{n} W_{nj}(T_i)X_j$. Liang 和 Li (2009) 定义如下的校正的惩罚最小二乘目标函数

$$\mathcal{L}(\Sigma_{uu}, \beta) = \frac{1}{2}\sum_{i=1}^{n}(\widetilde{Y}_i - \widetilde{X}_i^{\mathrm{T}}\beta)^2 - \frac{n}{2}\beta^{\mathrm{T}}\Sigma_{uu}\beta + n\sum_{j=1}^{p} p_{\lambda_j}(|\beta_j|), \tag{5.4.1}$$

其中 $p_{\lambda_j}(\cdot)$ 可取任何一个惩罚函数, λ_j 是截断参数, 用于控制参数估计的稀疏性, 可以通过一些数据驱动的准则进行选取. 在假定 Σ_{uu} 已知时, 关于参数 β 极小化校正的惩罚最小二乘目标函数 (5.4.1), 则可得到参数 β 的惩罚最小二乘估计 $\hat{\beta}$.

校正的惩罚最小二乘目标函数 (5.4.1) 是针对部分线性 EV 模型 (5.2.1) 完成变量选择的一个一般框架. 如果惩罚函数 $p_{\lambda_j}(\cdot)$ 取为 L_0 惩罚, 即 $p_{\lambda_j}(|\beta_j|) = 0.5\lambda_j^2 I(|\beta_j| \neq 0)$, 其中 $I(\cdot)$ 是示性函数. L_0 惩罚函数的变量选择包含了 AIC (Akaike, 1973), BIC (Schwarz, 1978), RIC (Foster and George, 1994) 等传统的变量选择方法, 即考虑下面的惩罚最小二乘目标函数

$$\mathcal{L}(\Sigma_{uu}, \beta) = \frac{1}{2}\sum_{i=1}^{n}(\widetilde{Y}_i - \widetilde{X}_i^{\mathrm{T}}\beta)^2 - \frac{n}{2}\beta^{\mathrm{T}}\Sigma_{uu}\beta + \frac{n}{2}\sum_{j=1}^{p}\lambda_j^2 I(|\beta_j| \neq 0), \tag{5.4.2}$$

其中 $\sum_{j=1}^{p} I(|\beta_j| \neq 0)$ 等于被选模型的大小. 特别地, AIC, BIC 和 RIC 分别对应的取 $\lambda_j = \sigma\sqrt{2/n}$, $\sigma\sqrt{\log(n)/n}$ 和 $\sigma\sqrt{\log(p)/n}$. 注意到 L_0 惩罚不是连续的, 因此优化 (5.4.2) 要求搜索超过 2^p 个子集, 计算量非常大, 而且也很不稳定, 这也是传统变量选择方法的主要缺点. 统计学者为了克服传统变量选择方法的缺点, 提出使用连续和光滑的惩罚函数, 如桥回归 (Frank and Friedman, 1993), 取 L_q 惩罚 $p_\lambda(|\beta_j|) = \lambda|\beta_j|^q$; Lasso 惩罚 (Tibshirani, 1996), 即取 L_1 惩罚 $p_\lambda(|\beta_j|) = \lambda|\beta_j|$; SCAD 惩罚 (Fan and Li, 2001), 定义为

$$p_\lambda(|\beta|) = \begin{cases} \lambda|\beta|, & 0 \leqslant |\beta| < \lambda, \\ \dfrac{(a^2-1)\lambda^2 - (|\beta| - a\lambda)^2}{2(a-1)}, & \lambda \leqslant |\beta| < a\lambda, \\ \dfrac{(a+1)\lambda^2}{2}, & |\beta| \geqslant a\lambda, \end{cases} \tag{5.4.3}$$

其中 Fan 和 Li (2001) 建议取 $a = 3.7$. SCAD 惩罚函数具有良好的性质, 满足: ① 稀疏性; ② 无偏性; ③ 连续性. 为了给出校正惩罚最小二乘估计的渐近性质, 首先介绍一些概念. 令 $\beta_0 = (\beta_{10}, \cdots, \beta_{p0})^{\mathrm{T}} = (\beta_{I0}^{\mathrm{T}}, \beta_{II0}^{\mathrm{T}})^{\mathrm{T}}$ 为 β 的真值, 其中 β_{I0} 表示 s 维的非零参数向量, $\beta_{II0} = 0$. 令

$$a_n = \max_{1 \leqslant j \leqslant p} \left\{ |p'_{\lambda_j}(|\beta_{j0}|)|, \beta_{j0} \neq 0 \right\}, \quad b_n = \max_{1 \leqslant j \leqslant p} \left\{ |p''_{\lambda_j}(|\beta_{j0}|)|, \beta_{j0} \neq 0 \right\},$$

$$b = \left(p'_{\lambda_1}(|\beta_{10}|)\mathrm{sgn}(\beta_{10}), \cdots, p'_{\lambda_s}(|\beta_{s0}|)\mathrm{sgn}(\beta_{s0}) \right)^{\mathrm{T}}$$

和 $\Sigma_\lambda = \mathrm{diag}\{p''_{\lambda_1}(|\beta_{10}|), \cdots, p''_{\lambda_s}(|\beta_{s0}|)\}$. 为了证明估计量的渐近结果, 首先给出下面的正则条件.

(C5.4.1) 假设 $\Sigma_{x|T} = \mathrm{Cov}\{x_I - E(x_I|T)\}$ 为正定矩阵, 其中 x_I 表示非零系数所对应的协变量构成的向量. $E(\varepsilon|x, T) = 0$ 和 $E(|\varepsilon|^3|x, T) < \infty$.

(C5.4.2) 估计非参数函数的窗宽取 $h = h_0 n^{-1/5}$, 其中 $h_0 > 0$ 为某一常数.

(C5.4.3) $K(\cdot)$ 是一个具有有界支撑的有界对称密度函数, 并满足

$$\int K(u)\mathrm{d}u = 1, \quad \int uK(u)\mathrm{d}u = 0, \quad \int u^2 K(u)\mathrm{d}u = 1.$$

(C5.4.4) T 在 $[0,1]$ 上有连续密度函数 $f(t)$, 有有界的连续二阶导数, 且 $0 < \inf\limits_{0\leqslant t\leqslant 1} f(t) \leqslant \sup\limits_{0\leqslant t\leqslant 1} f(t) < \infty$. 此外, (Y, T) 的密度函数有界远离 0, 并有有界的连续二阶导数.

(C5.4.5) $g_Y(t) = E(Y|T = t)$ 和 $g_X(t) = E(X|T = t)$ 是有界的, 并有连续的二阶导数.

下面给出 Liang 和 Li (2009) 中估计的渐近结果, 定理的详细证明请参考 Liang 和 Li (2009).

定理 5.4.1 假设 $a_n = O(n^{-1/2}), b_n \to 0$, 且正则条件(C5.4.1)～条件(C5.4.5)成立, 则有下面的结论:

(1) 依概率趋于 1, $\mathcal{L}(\Sigma_{uu}, \beta)$ 存在一个局部极小值 $\hat{\beta}$, 使得 $\|\hat{\beta} - \beta\| = O_P(n^{-1/2})$;

(2) 进一步假设所有的 $\lambda_j \to 0, n^{1/2}\lambda_j \to \infty$, 和

$$\liminf_{n\to\infty} \{\liminf_{v\to 0^+} p'_{\lambda_j}(v)/\lambda_j\} > 0. \tag{5.4.4}$$

依概率趋于 1, 在结论 (1) 中的 $\sqrt{n}$ 相合估计 $\hat{\beta}$ 满足: (i) $\hat{\beta}_{II} = 0$; (ii) $\hat{\beta}_I$ 有如下的渐近正态分布

$$\sqrt{n}(\Sigma_{x|T} + \Sigma_\lambda)\{\hat{\beta}_I - \beta_{I0} + (\Sigma_{x|T} + \Sigma_\lambda)^{-1} b\} \xrightarrow{L} N(0, \Gamma^*),$$

其中 $\Gamma^* = E\left[(\varepsilon - u_I^{\mathrm{T}}\beta_{I0})\{x_I - E(x_I|T)\} + (u_I u_I^{\mathrm{T}} - \Sigma_{uu,I})\beta_{I0} + \varepsilon u_I\right]^{\otimes 2}$, u_I 和 $\Sigma_{uu,I}$ 分别表示测量误差 u 和 Σ_{uu} 对应的非零系数对应的误差变量和协方差矩阵.

如果使用定理 5.4.1 对 β_{I0} 进行统计推断, 需要得到 $\hat{\beta}_I$ 的渐近协方差矩阵的相合估计. 由定理 5.4.1, $\hat{\beta}_I$ 的渐近协方差矩阵为

$$\frac{1}{n}(\Sigma_{x|T}+\Sigma_\lambda)^{-1}\Gamma^*(\Sigma_{x|T}+\Sigma_\lambda)^{-1}.$$

首先 $\Sigma_{x|T}$ 的相合估计定义为

$$\hat{\Sigma}_{x|T}=\frac{1}{n-s}\sum_{i=1}^{n}\widetilde{X}_{i,I}^{\otimes 2}-\Sigma_{uu,I}.$$

Γ^* 的相合估计为

$$\hat{\Gamma}^*=\frac{1}{n}\sum_{i=1}^{n}\left\{\widetilde{X}_{i,I}(\widetilde{Y}_i-\widetilde{X}_{i,I}^{\mathrm{T}}\hat{\beta}_I)+\Sigma_{uu,I}\hat{\beta}_I\right\}^{\otimes 2}.$$

由 plug-in 方法可得 $\hat{\beta}_I$ 的渐近协方差矩阵的相合估计为

$$\frac{1}{n}\left\{\hat{\Sigma}_{x|T}+\Sigma_\lambda(\hat{\beta}_I)\right\}^{-1}\hat{\Gamma}^*\left\{\hat{\Sigma}_{x|T}+\Sigma_\lambda(\hat{\beta}_I)\right\}^{-1},$$

其中 $\Sigma_\lambda(\hat{\beta}_I)$ 是把 Σ_λ 中的 β_I 用其估计量 $\hat{\beta}_I$ 替换而得到的.

在许多实际问题中, Σ_{uu} 是未知的, 因此需要估计 Σ_{uu} 的大小. Liang 等 (1999) 建议采用重复测量的方法来估计测量误差的方差阵 Σ_{uu}. 假定重复测量, 即 $X_{ij}=x_i+u_{ij}, j=1,\cdots,m_i$. 令 $\bar{X}_i=\dfrac{1}{m_i}\sum_{j=1}^{m_i}X_{ij}$ 表示重复测量的样本均值, 则 Σ_{uu} 的一个无偏的矩估计 $\hat{\Sigma}_{uu}$ 可以由式 (5.2.6) 进行定义. 令 $\bar{u}_i=m_i^{-1}\sum_{j=1}^{m_i}u_{ij}$, $\mathrm{Cov}(\bar{u}_i)=m_i^{-1}\Sigma_{uu}$, 则可定义校正的惩罚最小二乘目标函数为

$$\mathcal{L}(\hat{\Sigma}_{uu},\beta)=\frac{1}{2}\sum_{i=1}^{n}\left\{\left(\widetilde{Y}_i-\widetilde{\bar{X}}_i^{\mathrm{T}}\beta\right)^2-m_i^{-1}\beta^{\mathrm{T}}\hat{\Sigma}_{uu}\beta\right\}+n\sum_{j=1}^{p}p_{\lambda_j}(|\beta_j|),\tag{5.4.5}$$

其中 $\widetilde{\bar{X}}_i=\bar{X}_i-\hat{g}_{\bar{W}}(T_i)$, 且 $\hat{g}_{\bar{W}}(t)$ 是 $E(\bar{X}_i|T_i=t)$ 的核估计或局部线性估计. 关于 β 极小化式 (5.4.5), 可得 β 的惩罚最小二乘估计, Liang 和 Li (2009) 证明该惩罚最小二乘估计的相合性和 oracle 性质.

5.4.2　惩罚分位数回归

当数据中存在离群点的时候, 最小二乘方法不能有效完成参数估计, 这时一些稳健的估计方法被提出来有效的完成参数估计等问题, 如 M 估计方法 (Huber, 1981) 和分位数回归 (Koenker, 2005). 对于没有测量误差的线性模型, Li 等 (2001) 提出了稳健的非凹惩罚的 M 估计方法进行变量选择研究. 对于模型 (5.2.1), Liang

和 Li (2009) 提出了稳健的变量选择方法——惩罚分位数回归, 下面就这一方法进行介绍.

对具有离群点的测量误差模型, 如果不考虑离群点的影响, 直接用上面介绍的校正的惩罚最小二乘估计方法进行变量选择时, 会导致得到的参数估计表现很差, 另外也会导致变量选择方法失效. 但是如果考虑离群点的影响, 使用稳健的分位数回归方法对参数进行估计时, 遇到的一个主要问题是如何进行校正衰减来消除测量误差对参数估计不相合性的影响. 在校正的惩罚最小二乘目标函数中, 可以使用项 $-n\beta^{\mathrm{T}}\Sigma_{uu}\beta$ 来校正估计的偏差, 产生参数 β 的相合估计. 但是这种方法不能直接推广到惩罚分位数回归来减少估计的偏差, He 和 Liang (2000) 对线性测量误差模型和部分线性测量误差模型应用正交回归的想法, 解决了测量误差对分位数回归方法的影响, 证明了估计量是相合的和渐近正态的. Liang 和 Li (2009) 把正交回归的思想推广到模型 (5.2.1) 的研究, 提出了惩罚分位数回归方法. 正交回归的核心思想是, 代替经典的残差平方和, 目标函数是通过从数据点到回归函数直线的正交距离的平方和来定义.

为了定义测量误差数据的分位数回归的正交回归方法, 假设随机向量 $(\varepsilon, u^{\mathrm{T}})$ 来自于均值为 0, 协方差矩阵为 $\sigma^2\Sigma$ 的椭球分布, 其中 σ^2 是未知的, 而 Σ 是一个分块矩阵, 第 (1,1) 位置的元素为 1, 剩下的 $p\times p$ 的对角矩阵为 C_{uu}. 进一步假定矩阵 C_{uu} 正比于误差协方差阵 Σ_{uu}, 并假定是已知的. 令 ρ_τ 表示第 τ 个分位数目标函数, 定义为

$$\rho_\tau(r) = \tau\max(r,0) + (1-\tau)\max(-r,0). \tag{5.4.6}$$

注意到关于 $v\in\mathbb{R}$ 极小化 $E\rho_\tau(\varepsilon - v)$ 的解是 ε 的第 τ 个分位数. 定义下面的惩罚分位数目标函数:

$$\mathcal{L}_\tau(\beta) = \sum_{i=1}^{n}\rho_\tau\left(\frac{\widetilde{Y}_i - \widetilde{X}_i^{\mathrm{T}}\beta}{\sqrt{1+\beta^{\mathrm{T}}C_{uu}\beta}}\right) + n\sum_{j=1}^{p}p_{\lambda_j}(|\beta_j|). \tag{5.4.7}$$

对比 He 和 Liang (2000) 提出的分位数回归估计方法, 式 (5.4.7) 右边增加了一个惩罚项, 该惩罚项主要用于产生参数 β 的稀疏解. 此外, He 和 Liang (2000) 假设 $(\varepsilon, u^{\mathrm{T}})$ 的分布来自于一个球形分布, 即 $C_{uu} = I_p$, 也就是假设了测量误差向量 u 的元素之间是不相关的. Liang 和 Li (2009) 放松了这一假设, 假设 $(\varepsilon, u^{\mathrm{T}})$ 的分布来自于一个椭球分布.

不同于校正的惩罚最小二乘目标函数 (5.4.1), 在惩罚分位数目标函数中使用因子 $\sqrt{1+\beta^{\mathrm{T}}C_{uu}\beta}$ 来修正测量误差对目标损失函数所引起的偏差. 为了更好的理解该思想, 下面介绍一个关于椭球分布的基本结论.

令 ω 表示一个随机向量, 并来自一个椭球分布, 其特征函数表示为

$$E\{\exp(\mathrm{i}t^{\mathrm{T}}\omega)\} = \exp(\mathrm{i}\mu^{\mathrm{T}}t)\phi(t^{\mathrm{T}}\Lambda t).$$

由 Fang 等 (1990), 可知 $E(\omega) = \mu$ 和 $\mathrm{Cov}(\omega) = \{-2\phi'(0)\}\Lambda$. 进一步, 如果 $\mu = 0$, 对于任给的常数向量 $b \neq 0$, 则有

$$\omega_1 \overset{d}{=} \frac{b^{\mathrm{T}}\omega}{\sqrt{b^{\mathrm{T}}\Lambda b/\Lambda_{11}}}, \tag{5.4.8}$$

其中 ω_1 是随机向量 ω 的第一个元素, Λ_{11} 是矩阵 Λ 的第 (1,1) 个元素, $R_1 \overset{d}{=} R_2$ 表示 R_1 与 R_2 有相同的分布.

由一些简单计算和式 (5.4.8) 的结论, 容易证明 ε_i 与 $(\varepsilon_i - u_i^{\mathrm{T}}\beta)/\sqrt{1+\beta^{\mathrm{T}}C_{uu}\beta}$ 有相同的分布. 因此, 在惩罚分位数目标函数 (5.4.7) 中的损失函数自动消除了测量误差所引起的偏差.

定理 5.4.2　假设 $a_n = O(n^{-1/2}), b_n \to 0$, 随机向量 $(\varepsilon_i, u_i^{\mathrm{T}})$ 来自于均值为 0, 协方差矩阵为 $\sigma^2\Sigma$ 的椭球分布, $i = 1, \cdots, n$. 进一步假设方程 $E\rho_\tau(\varepsilon_i - q) = 0$ 有唯一的解, 表示为 q_τ, 且当 $\vartheta \to 0$ 时, $f_\varepsilon(\vartheta + q_\tau) - f_\varepsilon(q_\tau) = O(\vartheta^{1/2})$, 其中 $f_\varepsilon(\cdot)$ 表示 ε 的密度函数, 则有下面的结论:

(1) 依概率趋于 1, $\mathcal{L}_\tau(\beta)$ 存在一个局部极小值 $\hat{\beta}_\tau$, 使得 $\|\hat{\beta}_\tau - \beta\| = O_P(n^{-1/2})$;

(2) 进一步假设所有的 $\lambda_j \to 0, n^{1/2}\lambda_j \to \infty$, 和式 (5.4.4) 成立. 依概率趋于 1, 在结论 (1) 中的 $\sqrt{n}$ 相合估计 $\hat{\beta}_\tau = (\hat{\beta}_{I,\tau}^{\mathrm{T}}, \hat{\beta}_{II,\tau}^{\mathrm{T}})^{\mathrm{T}}$ 满足: (i) $\hat{\beta}_{II,\tau} = 0$; (ii) $\hat{\beta}_{I,\tau}$ 有如下的渐近正态分布

$$\sqrt{n}f_\varepsilon(q_\tau)(1+\beta_{I0}^{\mathrm{T}}C_{11}\beta_{I0})^{1/2}\Sigma_{x|T}\{\hat{\beta}_{I,\tau} - \beta_{I0}\} + nb(1+\beta_{I0}^{\mathrm{T}}C_{11}\beta_{I0})^{1/2} \xrightarrow{L} N(0, J_q),$$

其中 C_{11} 是矩阵 C_{uu} 的前 $s \times s$ 个子对角矩阵, 且

$$J_q = \tau(1-\tau)\Sigma_{x|T} + \mathrm{Cov}\left\{\psi_\tau(\xi)\left(u_{11} + \frac{\xi\beta_{I0}}{\sqrt{1+\beta_{I0}^{\mathrm{T}}C_{11}\beta_{I0}}}\right)\right\},$$

$\xi = (\varepsilon - u_{11}^{\mathrm{T}}\beta_{I0})\Big/\sqrt{1+\beta_{I0}^{\mathrm{T}}C_{11}\beta_{I0}} - q_\tau$ 和 ψ_τ 表示 ρ_τ 的导数.

特别值得注意的是, 当 $C_{uu} = I_p$ 和 $b = 0$ 时, 相应于式 (5.4.7) 中没有惩罚项, 则定理 5.4.2 的渐近正态性结果相同于 He 和 Liang (2000) 的结果.

5.4.3　估计方法的实施

在上面两种惩罚变量选择研究中会涉及三个实际需要解决的问题: 光滑参数窗宽 h 的选取, 截断参数 $\lambda = (\lambda_1, \cdots, \lambda_p)$ 的选取, 极小化校正的惩罚最小二乘和惩罚分位数目标函数的算法问题.

1. 窗宽 h 的选取

条件 (C5.4.2) 中要求取 $h = h_0 n^{-1/5}$, $h_0 > 0$ 为常数, 而且利用这个窗宽估计未知函数不会影响参数估计量的渐近性质. 因此, Liang 和 Li (2009) 建议采用 Ruppert 等 (1995) 提出的 plug-in 窗宽选取准则选取窗宽, 并使选取的窗宽满足条件 (C5.4.2). 在 Liang 和 Li (2009) 的模拟研究中, 显示该窗宽选取准则是稳定的.

2. 局部二次逼近

极小化校正的惩罚最小二乘目标函数 (5.4.1) 或惩罚分位数目标函数 (5.4.7) 求解 β 的估计是一个比较困难的问题, 因为惩罚函数在原点处奇异并不存在连续的二阶导数, 并且关于 β 是非凹的. 这种奇异性和非凹性对优化目标函数提出了一定的挑战. 为了解决这个问题, Fan 和 Li (2001) 建议使用二次函数局部逼近惩罚函数, 提出一种局部二次逼近 (local quadratic approximation, LQA) 的算法. 基于 LQA 算法和调整的 Newton-Raphson 迭代算法, 可以完成惩罚目标函数的最优化问题, 完成 β 的估计. Fan 和 Li (2001), 与 Hunter 和 Li (2005) 指出, LQA 算法分享了向后逐步回归变量选择方法的缺点: 在 LQA 算法中, 某一迭代步骤中一个协变量被删除, 它必将永远不会出现在最终的模型中, 如果该删除变量是显著性的, 将直接导致选取的模型是错误的. 为了克服该算法问题, Hunter 和 Li (2005) 提出了一个 MM (minorize-maximize) 算法, 尽管 MM 算法可以解决 LQA 算法的缺陷, 但是需要选取一个扰动参数, 在实践中也有一定的难度. 因此, 也把 Hunter 和 Li (2005) 提出的 MM 算法称为扰动的 LQA 算法. Zou 和 Li (2008) 进一步提出了一个修正 LQA 算法的局部线性逼近 (local linear approximation, LLA) 算法. Liang 和 Li (2009) 使用扰动的 LQA 算法来完成两种惩罚变量选择, 下面对扰动的 LQA 算法进行介绍.

对于给定的初始值 $\beta^{(0)}$, 其接近于真实参数向量 β. 如果 β_j 接近于 0, 则设 $\hat{\beta}_j = 0$. 否则对于 $\beta_j \approx \beta_j^{(0)}$, 惩罚函数可以用一个二次函数局部逼近如下

$$p_{\lambda_j}(|\beta_j|) \approx p_{\lambda_j}(|\beta_j^{(0)}|) + \frac{1}{2}\left\{\frac{p'_{\lambda_j}(|\beta_j^{(0)}|)}{|\beta_j^{(0)}| + \tau_0}\right\}(\beta_j^2 - \beta_j^{(0)2}) =: q_{\lambda_j}(|\beta_j|), \tag{5.4.9}$$

其中 τ_0 为指定的扰动参数. τ_0 的大小影响着解的稀疏性和收敛速度, 在实际应用中很难指定一个合适的大小. Hunter 和 Li (2005) 建议使用

$$\tau_0 = \frac{\xi}{2n\lambda_j} \min\left\{|\beta_j^{(0)}| : \beta_j^{(0)} \neq 0\right\}, \tag{5.4.10}$$

其中 ξ 为给定的容忍参数, 更多的讨论见 Hunter 和 Li (2005).

基于式 (5.4.9) 的局部二次逼近, 调整的 Newton-Raphson 迭代算法能够用于寻找校正的惩罚最小二乘目标函数或惩罚分位数目标函数的最小解. 如果初始值

$\beta^{(0)}$ 取没有惩罚的估计值, 并去掉常数项, 则可以用下面的迭代公式求解. 对于校正的惩罚最小二乘目标函数, 当 Σ_{uu} 已知时, 迭代公式为

$$\beta^{(k+1)} = \arg\min_{\beta} \left\{ \frac{1}{2} \sum_{i=1}^{n} (\widetilde{Y}_i - \widetilde{X}_i^{\mathrm{T}} \beta)^2 - \frac{n}{2} \beta^{\mathrm{T}} \Sigma_{uu} \beta + n \sum_{j=1}^{p} \frac{p'_{\lambda_j}(|\beta_j^{(k)}|)}{2(|\beta_j^{(k)}| + \tau_0)} \beta_j^2 \right\}.$$

当 Σ_{uu} 未知时, 迭代公式为

$$\beta^{(k+1)} = \arg\min_{\beta} \left[\frac{1}{2} \sum_{i=1}^{n} \left\{ \left(\widetilde{Y}_i - \widetilde{\widetilde{X}}_i^{\mathrm{T}} \beta \right)^2 - \frac{1}{m_i} \beta^{\mathrm{T}} \hat{\Sigma}_{uu} \beta \right\} + n \sum_{j=1}^{p} \frac{p'_{\lambda_j}(|\beta_j^{(k)}|)}{2(|\beta_j^{(k)}| + \tau_0)} \beta_j^2 \right].$$

对于惩罚分位数目标函数, 迭代公式为

$$\beta^{(k+1)} = \arg\min_{\beta} \left\{ \sum_{i=1}^{n} \rho_\tau \left(\frac{\widetilde{Y}_i - \widetilde{X}_i^{\mathrm{T}} \beta}{\sqrt{1 + \beta^{\mathrm{T}} C_{uu} \beta}} \right) + n \sum_{j=1}^{p} \frac{p'_{\lambda_j}(|\beta_j^{(k)}|)}{2(|\beta_j^{(k)}| + \tau_0)} \beta_j^2 \right\}.$$

对于给定的截断参数 λ, 反复迭代上面三个公式, 直到收敛, 即可求得 β 的最终的惩罚最小二乘估计和惩罚分位数估计.

3. 截断参数 λ 的选取

此处以校正的惩罚最小二乘为例, 介绍截断参数 $\lambda = (\lambda_1, \cdots, \lambda_p)^{\mathrm{T}}$ 的选取, 对于惩罚分位数是类似的. 在数据分析中, 可以极小化广义交错核实 (GCV) 准则选取截断参数 λ. 但是 Wang 等 (2007) 证明, 即使样本量 n 趋于无穷大, GCV 选取准则有过分拟合效应, 也不具有模型选择的相合性. Wang 等 (2007) 提出了一种基于 BIC 的截断参数的选取准则, 并证明该方法具有模型选择的相合性, 即可以依概率 1 正确识别真实模型. 基于此, Liang 和 Li (2009) 采用 BIC 准则选取截断参数 λ. 令

$$l(\beta) = \frac{1}{2} \sum_{i=1}^{n} \left\{ \left(\widetilde{Y}_i - \widetilde{\widetilde{X}}_i^{\mathrm{T}} \beta \right)^2 - {m_i}^{-1} \beta^{\mathrm{T}} \hat{\Sigma}_{uu} \beta \right\}. \tag{5.4.11}$$

由式 (5.4.5) 和式 (5.4.11), 类似于 Fan 和 Li (2001), 定义有效的参数个数为

$$e(\lambda) = \mathrm{tr}\left[\{\mathcal{L}''(\hat{\Sigma}_{uu}, \hat{\beta})\}^{-1} l''(\hat{\beta}) \right],$$

其中 $\lambda = (\lambda_1, \cdots, \lambda_p)^{\mathrm{T}}$, 对于任意的函数 $G(\beta)$, $G''(\beta) = \partial G(\beta)/\partial\beta\beta^{\mathrm{T}}$ 表示函数 $G(\beta)$ 的 Hessian 矩阵. Liang 和 Li (2009) 定义了如下的 BIC 信息准则:

$$\mathrm{BIC}(\lambda) = \log(\mathrm{RSS}_\lambda) + e(\lambda)\log(n)/n, \tag{5.4.12}$$

其中 RSS_λ 表示基于截断参数 λ 由惩罚最小二乘方法所得到的残差平方和. 极小化 $\mathrm{BIC}(\lambda)$ 可获得截断参数的大小, 但是实际使用时会遇到一个问题, 即需要在一个 p 维空间极小化 $\mathrm{BIC}(\lambda)$, 当 p 很大时, 难度非常大. 为了解决这个问题, Fan 和 Li (2001) 认为 λ_j 的大小是与 β_j 的没有惩罚的最小二乘估计 $\hat{\beta}_j^u$ 的标准误差是成比例的, 这样取 $\lambda_j = \lambda^* \times \mathrm{SE}(\hat{\beta}_j^u)$, 其中 $\mathrm{SE}(\hat{\beta}_j^u)$ 表示没有惩罚的最小二乘估计 $\hat{\beta}_j^u$ 的标准误差. 这时就把在 p 维空间上的极小化问题转化成了在 1 维空间上的极小化问题. 只需要极小化 $\mathrm{BIC}(\lambda^*)$, 选取一个 $\hat{\lambda}^*$, 即可获得相应的截断参数 $\hat{\lambda}_j = \hat{\lambda}^*\mathrm{SE}(\hat{\beta}_j^u), j = 1, \cdots, p$. 更多详细的讨论见 Fan 和 Li (2001), Wang 等 (2007), Liang 和 Li (2009), 与 Li 等 (2011b).

5.5 协变量调整部分线性模型的变量选择

本节主要考虑协变量调整部分线性模型的变量选择问题, 所提出的惩罚最小二乘估计方法能够同时进行变量选择和参数估计. 在一些正则条件下, 证明估计量的收敛速度和渐近正态性. 在适当的惩罚函数和截断参数下, 所提出的程序完成和 oracle 估计有相同的效率.

5.5.1 模型介绍

本节考虑如下的协变量调整半参数部分线性模型:

$$\begin{cases} Y = X^{\mathrm{T}}\beta_0 + g(T) + \varepsilon, \\ \widetilde{Y} = \phi(U)Y, \\ \widetilde{X}_r = \psi_r(U)X_r, \quad r = 1, \cdots, p, \end{cases} \tag{5.5.1}$$

其中 Y 是不可观测的响应变量, $X = (X_1, X_2, \cdots, X_p)^{\mathrm{T}}$ 是不可观测的连续型预测向量, $\widetilde{Y}$ 和 $\widetilde{X}_r$ 是测量到的混淆变量, $\phi(U)$ 和 $\psi_r(U)$ 是变量 U 的未知的失真函数, U 是一个有界的观测协变量, $g(\cdot)$ 是一个未知的光滑函数, 且 β_0 是未知的回归系数向量. 不失一般性, 假定变量 T 是区间 [0,1] 上的一元变量. 当然所提方法可以推广到多元变量 T, 但在实际使用中会遇到维数祸根问题. 在真实模型中, 假设 β_0 中的一些分量是零. 本节的主要目的就是依概率 1 识别这些零系数, 并且估计非零回归系数和非参数函数 $g(\cdot)$.

为了介绍估计程序, 首先引入如下一些关于模型 (5.5.1) 基本的假设条件.

(1) $E[\phi(U)] = 1, E[\psi_r(U)] = 1, r = 1, \cdots, p$.

(2) 对给定的 $T = t$, $E[X_r|T = t] = \theta_r(t)$, 即 $X_r = \theta_r(T) + \eta_r$. $g(\cdot)$ 和 $\theta_r(t)$ 的前两阶导数是一阶 Lipschitz 连续的, $r = 1, \cdots, p$.

(3) (X, U, ε) 是相互独立的, $(U, T, \varepsilon, \eta)$ 也是相互独立的, 其中 $\eta = (\eta_1, \cdots, \eta_p)^{\mathrm{T}}$, 且 (Y, U) 也是相互独立的.

(4) $E(\varepsilon)=0, \mathrm{Var}(\varepsilon)=\sigma^2$ 和 $E(\eta)=0, \mathrm{Var}(\eta)=\Sigma$.

(5) $\sup_t E(\|X\|^4|T=t)<\infty$ 和 $E(\varepsilon^4+U^4)<\infty$.

上面的条件是一般的, 并在大部分情况下是可以满足的. 条件 (1) 是一个一般的假设, 它是一个识别性条件, 被共同用于协变量调整模型, 见 Sentürk 和 Müller (2005a, 2005b), Cui (2008), Cui 等 (2009), Zhang 等 (2012a, 2012b), Zhang (2013a, 2013b). 条件 (2) 常用于半参数模型的研究, 见 He 等 (2002), Leng 等 (2010), 这个条件假设了变量 X 和 T 之间的相依性, 并保证了非参数 $g(\cdot)$ 是充分光滑的, 以及决定了其估计量的收敛速度. 条件 (3) 用于部分线性协变量调整模型中, 见 Li 等 (2010a). 条件 (4) 控制了 β_0 估计量方差的渐近行为. 条件 (5) 是一个技术条件, 用于推导估计量的渐近性质.

最近, 收缩的惩罚变量选择方法已经被成功推广到了半参数回归模型的研究中, 例如, Ni 等 (2009), Xie 和 Huang (2009), Li 和 Liang (2008), Liang 和 Li (2009), Liu 等 (2011). 然而在许多实际应用中, 响应变量和协变量不能被直接观测, 而观测到的变量是受到失真函数污染以后的测量误差数据, 如 Sentürk 和 Müller (2005a, 2005b), Sentürk (2006), Cui 等 (2009), Zhang 等 (2012a, 2012b), Zhang 等 (2013a, 2013b), Zhang 等 (2014) 等学者已经对该协变量调整回归模型进行了研究, 然而这些收缩的惩罚变量选择方法不能直接用于测量误差模型或协变量调整回归模型. 对于包含两个或两个以上协变量的线性模型, 其中一些协变量可能被失真函数或测量误差污染, 而其余变量是真实观测变量, 这时可以证明如果使用观测变量代替不可观测的真实变量得到的回归系数的最小二乘估计是不相合的 (Cui et al., 2009). 实践中, 收集到的初始数据需要进行评价, 即分析变量是否是真实被测量, 以及使用何种模型有效地去分析或拟合数据. 在数据分析中, 也要确定哪些协变量需要保留在模型中, 而哪些协变量是没有信息的. 例如, Sentürk (2006) 使用协变量调整变系数模型去分析肌酐数据 (creatinine data), 并建议体表面积 (BSA) 作为混淆变量, Li 等 (2010a) 进一步分析该数据发现协变量调整部分线性模型能够更有效地拟合该数据. 但是 Li 等 (2010a) 仅仅考虑了钙的摄入量关于钙吸收为线性效应, 年龄为非线性效应. 在本节的实例分析部分, 发现年龄对钙吸收有一个很强的非线性效应 (为一个 U 型曲线), 而变量体重指数 (body mass index, BMI)、膳食钙 (Cald)、尿钙 (Calu)、内源性粪 (Efc) 和磷 (Pho) 对响应变量钙吸收有线性效应. 本节的主要目的就是进一步发展协变量调整半参数部分线性模型, 并用该模型分析肌酐数据, 探索哪些变量对于钙吸收是显著的. 由于该数据存在失真的测量误差, 存在的收缩的惩罚变量选择方法不能直接用于该模型的研究, 如果直接用观测变量进行研究, 就不能有效地找到究竟是哪些重要变量对钙吸收有显著影响, 即变量选择方法便无效. 因此, 需要进一步发展协变量调整部分线性模型的变量选择方法, 理论上系统研究变量选择估计的渐近性质.

5.5.2 变量选择方法

假设数据 $\{(U_i,T_i,\widetilde{X}_i,\widetilde{Y}_i),i=1,\cdots,n\}$ 是 i.i.d. 的, 并来自于模型 (5.5.1), 其中 $\widetilde{X}_i=(\widetilde{X}_{1i},\cdots,\widetilde{X}_{pi})^{\mathrm{T}}$ 是 p 维可观测的协变量. 类似于 Cui 等 (2009), 有

$$\phi(U)=\frac{E[\widetilde{Y}|U]}{E[Y]},\quad \psi_r(U)=\frac{E[\widetilde{X}_r|U]}{E[X_r]},\quad r=1,\cdots,p.$$

由核光滑方法, 能够使用测量数据 $\{\widetilde{Y}_i,\widetilde{X}_i,U_i,1\leqslant i\leqslant n\}$ 估计 $\phi(U)$ 和 $\psi_r(U)(1\leqslant r\leqslant p)$ 为

$$\begin{aligned}\hat{\phi}(u)&=\frac{\dfrac{1}{n}\sum\limits_{i=1}^{n}K_h(u-U_i)\widetilde{Y}_i}{\dfrac{1}{n}\sum\limits_{i=1}^{n}K_h(u-U_i)}\times\frac{1}{\bar{\widetilde{Y}}},\\ \hat{\psi}_r(u)&=\frac{\dfrac{1}{n}\sum\limits_{i=1}^{n}K_h(u-U_i)\widetilde{X}_{ir}}{\dfrac{1}{n}\sum\limits_{i=1}^{n}K_h(u-U_i)}\times\frac{1}{\bar{\widetilde{X}}_r},\quad r=1,\cdots,p,\end{aligned}\tag{5.5.2}$$

其中 $\bar{\widetilde{Y}}=\frac{1}{n}\sum_{i=1}^{n}\widetilde{Y}_i$, $\bar{\widetilde{X}}_r=\frac{1}{n}\sum_{i=1}^{n}\widetilde{X}_{ir}$, $K_h(\cdot)=K(\cdot/h)/h$, $K(\cdot)$ 是核函数, h 是窗宽, $r=1,\cdots,p$. 随后, 分别得到 Y_i 和 X_{ir} 的估计为

$$\hat{Y}_i=\frac{\widetilde{Y}_i}{\hat{\phi}(U_i)},\quad \hat{X}_{ir}=\frac{\widetilde{X}_{ir}}{\hat{\psi}_r(U_i)},\quad r=1,\cdots,p,\quad i=1,\cdots,n.\tag{5.5.3}$$

为了获得 β_0 的稀疏估计, 定义如下的惩罚最小二乘目标函数

$$\sum_{i=1}^{n}\left(\hat{Y}_i-\hat{X}_i^{\mathrm{T}}\beta-g(T_i)\right)^2+n\sum_{j=1}^{p}p_{\lambda_j}(|\beta_j|),\tag{5.5.4}$$

其中 $p_{\lambda_j}(\cdot)$ 是具有截断参数 λ_j 的惩罚函数. 该截断参数影响着参数回归系数解的稀疏性, 能够使用交错核实法和 BIC 等数据驱动的准则进行选取, 详细的讨论见 Breiman (1996), Tibshirani (1996), Fan 和 Li (2001), Li 等 (2011b). 既然惩罚最小二乘目标函数 (5.5.4) 中包含未知函数 $g(t)$, 因此不能直接极小化式 (5.5.4) 获得参数回归系数的稀疏解. 注意到 $g(T)=E(Y|T)-E[X|T]^{\mathrm{T}}\beta_0$. 对于给定的 β, $g(\cdot)$ 的非参数估计为

$$\hat{g}(t,\beta)=\sum_{i=1}^{n}W_{ni}(t)(\hat{Y}_i-\hat{X}_i^{\mathrm{T}}\beta),\tag{5.5.5}$$

其中 $\{W_{ni}(t), 1 \leqslant i \leqslant n\}$ 是一族非负的权函数. 例如, 非负权函数可以取为

$$W_{nj}(t) = \frac{\bar{K}\left(\dfrac{t-T_j}{h_1}\right)}{\sum_{q=1}^{n} \bar{K}\left(\dfrac{t-T_q}{h_1}\right)}, \quad j = 1, \cdots, n, \tag{5.5.6}$$

其中 $\bar{K}(\cdot)$ 是一个概率密度函数, h_1 是趋于零的光滑参数, 称为窗宽. 对于 $1 \leqslant i \leqslant n$, 令 $\check{X}_i = \hat{X}_i - \sum_{j=1}^{n} W_{nj}(T_i)\hat{X}_i$ 和 $\check{Y}_i = \hat{Y}_i - \sum_{j=1}^{n} W_{nj}(T_i)\hat{Y}_j$. 把非参数 $g(\cdot)$ 的估计 (5.5.5) 代入到惩罚最小二乘目标函数 (5.5.4), 则有

$$Q_n(\beta) = \sum_{i=1}^{n} \left(\check{Y}_i - \check{X}_i^{\mathrm{T}}\beta\right)^2 + n\sum_{j=1}^{p} p_{\lambda_j}(|\beta_j|).$$

极小化目标函数 $Q_n(\beta)$, 可以得到 β_0 的惩罚最小二乘估计, 即

$$\hat{\beta} = \arg\min_{\beta} \ Q_n(\beta). \tag{5.5.7}$$

由式 (5.5.5) 和式 (5.5.7), $g(\cdot)$ 的估计定义为

$$\hat{g}(t) = \sum_{i=1}^{n} W_{ni}(t)(\hat{Y}_i - \hat{X}_i^{\mathrm{T}}\hat{\beta}).$$

5.5.3 渐近性质

为了给出估计量的渐近性质, 首先介绍一些概念. 不失一般性, 假设 β_0 分成两部分 $\beta_0 = (\beta_{I0}^{\mathrm{T}}, \beta_{II0}^{\mathrm{T}})^{\mathrm{T}}$, 其中 β_{I0} 表示参数向量 β_0 中的非零元素构成的向量, 而 $\beta_{II0} = 0$. 令 s 表示非零向量 β_{I0} 的维数. 进一步令 $Y = (Y_1, \cdots, Y_n)^{\mathrm{T}}$, $\mathbb{X}$ 表示 $n \times p$ 的矩阵. 相应于 β_0 的划分, 记 $\mathbb{X} = (\mathbb{X}_I, \mathbb{X}_{II})$, 其中 $\mathbb{X}_I$ 和 $\mathbb{X}_{II}$ 分别表示 $n \times s$ 和 $n \times (p-s)$ 的矩阵.

记

$$\Sigma_\lambda = \mathrm{diag}\{p''_{\lambda_1}(\beta_{01}), \cdots, p''_{\lambda_s}(\beta_{0s})\},$$
$$b_n = \{p'_{\lambda_1}(|\beta_{01}|)\mathrm{sgn}(\beta_{01}), \cdots, p'_{\lambda_s}(|\beta_{0s}|)\mathrm{sgn}(\beta_{0s})\}^{\mathrm{T}}$$

和

$$\mathbb{S} = \begin{pmatrix} W_{n1}(T_1) & W_{n2}(T_1) & \cdots & W_{nn}(T_1) \\ W_{n1}(T_2) & W_{n2}(T_2) & \cdots & W_{nn}(T_2) \\ \vdots & \vdots & & \vdots \\ W_{n1}(T_n) & W_{n2}(T_n) & \cdots & W_{nn}(T_n) \end{pmatrix},$$

其中 $W_{ni}(T_j)$ 定义在式 (5.5.6) 中. 进一步令 $\mathbb{B}$ 是 $s \times s$ 的矩阵, 其第 (r,k) 个元素为

$$\mathbb{B}(r,k) = \beta_r \beta_k E\left\{ \frac{EY^2}{(EY)^2}\mathrm{Var}(\phi(U)) + \frac{E(X_r X_k)}{EX_r EX_k}\mathrm{Cov}(\psi_k(U), \psi_r(U)) - \frac{EYX_k}{EYEX_k}\mathrm{Cov}(\psi_k(U), \phi(U)) - \frac{EYX_r}{EYEX_r}\mathrm{Cov}(\psi_r(U), \phi(U)) \right\}.$$

为了得到估计量 $\hat{\beta}$ 的渐近性质, 需要下面的正则条件.

(C5.5.1) 核函数 $K(\cdot)$ 在支撑集 $[-1,1]$ 上是对称、连续可微的密度函数, 满足

$$\int_{-1}^{1} K(u)\mathrm{d}u = 1, \quad \int_{-1}^{1} u^2 K(u)\mathrm{d}u \neq 0, \quad \int_{-1}^{1} |u|^i K(u)\mathrm{d}u < \infty, \quad i = 1,2,3.$$

(C5.5.2) 正实数序列窗宽 h_1, 满足: 当 $n \to \infty$ 时, $nh_1^4 \to 0$, $n^{1/2}h_1/\log^3 n \to \infty$. 权函数 $W_{nj}(\cdot)$ 满足

$$\sum_{j=1}^{n} W_{nj}(T_i) = 1, \quad i = 1, \cdots, n,$$

$$\max_{1 \leqslant i,j \leqslant n} W_{nj}(T_i) = O_p\left((nh_1)^{-1}\log n\right),$$

$$\max_{1 \leqslant i \leqslant n} \sum_{j=1}^{n} W_{nj}(T_i) I[\mid T_j - T_i \mid > h_1] = O_p(h_1).$$

(C5.5.3) 假设 U 的密度函数 $f(u)$, 和所有的函数 $g_r(u) = \psi_r(u)f(u), 1 \leqslant r \leqslant p$, $g_Y(u) = \phi(u)f(u)$, $\psi_r(u)$, $\phi(u)$ 都大于一个正常数, 并是可微的. 进一步假设它们的导数满足如下条件: 存在原点的一个邻域 (记为 Δ) 和一个常数 $c > 0$, 使得对于任意的 $\delta \in \Delta$, 有

$$|g_r^{(3)}(u+\delta) - g_r^{(3)}(u)| \leqslant c|\delta|, \ 1 \leqslant r \leqslant p,$$
$$|g_Y^{(3)}(u+\delta) - g_Y^{(3)}(u)| \leqslant c|\delta|,$$
$$|f^{(3)}(u+\delta) - f^{(3)}(u)| \leqslant c|\delta|.$$

(C5.5.4) 当 $n \to \infty$ 时, 窗宽 h 在 $O(n^{-1/4}\log n)$ 和 $O(n^{-1/8})$ 之间的范围内变化.

(C5.5.5) EY 和 EX_r 是有界远离零的, $1 \leqslant r \leqslant p$.

(C5.5.6) $|\widetilde{Y}_i|$ 和 $|\widetilde{X}_{ir}|$ 依概率 1 被常数 $0 < B < \infty$ 所界定, $1 \leqslant i \leqslant n, 1 \leqslant r \leqslant p$.

(C5.5.7) 存在常数 C_1 和 C_2, 使得

$$0 < C_1 < \lambda_{\min}(\Sigma) < \lambda_{\max}(\Sigma) < C_2,$$

其中 Σ 定义在假设 (4) 中, $\lambda_{\min}(\Sigma)$ 和 $\lambda_{\max}(\Sigma)$ 分别表示矩阵 Σ 的最小和最大特征值.

(C5.5.8) 假设矩阵 $(I-\mathbb{S})^{\mathrm{T}}(I-\mathbb{S})$ 的最大特征值是有界的.

(C5.5.9) 变量 T 有连续型的密度函数 $f_T(t)$, 满足

$$0<\inf_{0\leqslant t\leqslant 1} f_T(t)\leqslant \sup_{0\leqslant t\leqslant 1} f_T(t)<\infty.$$

条件 (C5.5.1) 常用于基于核估计的 $\sqrt{n}$ 相合性的渐近证明, 见 Zhu 和 Fang (1996), Li 等 (2010a). 条件 (C5.5.2) 常用于半参数回归模型的研究, 见 Liang 等 (1999), Shi 和 Lau (2000). 条件 (C5.5.3) 是假定了函数 $g_r(\cdot)$ 和 $g_Y(\cdot)$, 以及 U 的密度函数 $f(\cdot)$ 的光滑性. 条件 (C5.5.4) 给了估计未知的失真函数要求的窗宽的范围, 该条件包含了最优窗宽. 条件 (C5.5.5) 是一个技术条件, 主要避免估计量的分母过小而导致估计无效, 类似的假设条件见 Sentürk 和 Müller (2005a, 2005b). 条件 (C5.5.6) 的有界性假设主要是为了方便渐近结果的证明. 该条件可以放松到 $|\widetilde{Y}_i|$ 和 $|\widetilde{X}_{ir}|$ 的四阶矩有界, 但是会增大证明的复杂度. 条件 (C5.5.7)~条件 (C5.5.9) 是半参数部分线性模型中常用的条件, 见 Li 等 (2010a).

令 $a_n=\max\{|p'_{\lambda_j}(|\beta_{0j}|)|:\beta_{0j}\neq 0\}$, $b_n=\max\{|p''_{\lambda_j}(|\beta_{0j}|)|:\beta_{0j}\neq 0\}$, 其中 β_{0j} 是 β_0 的第 j 个分量. 此外, 还需要对惩罚函数施加下面的条件.

(A5.5.1) $\liminf\limits_{n\to\infty}\liminf\limits_{\theta\to 0+} p'_\lambda(\theta)/\lambda>0$.

(A5.5.2) $a_n=O(n^{-1/2})$.

(A5.5.3) 当 $n\to\infty$ 时, $b_n\to 0$.

下面的定理给出了主要的渐近结果, 包括 $\sqrt{n}$ 相合估计的存在性、估计的稀疏性和非零系数估计的渐近正态性.

定理 5.5.1 假设条件(C5.5.1)~条件(C5.5.9)成立, 同时惩罚函数 $p_\lambda(\cdot)$ 满足条件(A5.5.1)~条件(A5.5.3), 则有下面的结论.

(1) 存在惩罚最小二乘目标函数 $Q_n(\beta)$ 的一个局部解 $\hat{\beta}$, 使得 $\|\hat{\beta}-\beta_0\|=O_P(n^{-1/2})$.

(2) 当 $n\to\infty$ 时, 进一步假设 $\lambda_j\to 0, \sqrt{n}\lambda_j\to\infty$, 则有依概率趋于 1, $\sqrt{n}$ 相合的惩罚最小二乘估计 $\hat{\beta}=(\hat{\beta}_I^{\mathrm{T}},\hat{\beta}_{II}^{\mathrm{T}})^{\mathrm{T}}$ 满足

(i) (稀疏性) $\hat{\beta}_{II}=0$.

(ii) (渐近正态性)

$$\sqrt{n}\Sigma_I^{-1}(\Sigma_I+\Sigma_\lambda)\left[(\hat{\beta}_I-\beta_{I0})+\left(\Sigma_I+\Sigma_\lambda\right)^{-1}b_n\right]\xrightarrow{L}N(0,\mathbb{G}),$$

其中协方差矩阵 $\mathbb{G}$ 的定义为 $\mathbb{G}=\sigma^2\Sigma_I^{-1}+\mathbb{B}$, Σ_I 是矩阵 Σ 相应于 s 维非零系数的 $s\times s$ 的子矩阵.

下面的定理 5.5.2 证明了非参数估计 $\hat{g}(\cdot)$ 的最优收敛速度.

定理 5.5.2 在定理 5.5.1 的条件下, 有

$$\hat{g}(t) - g(t) = O_P(n^{-2/5}).$$

5.5.4 模拟研究和实际数据分析

1. 模拟研究

本节通过 Monte Carlo 模拟研究比较所提估计方法在有限样本下的完成情况. 在模拟研究中, 考虑两种样本量大小 $n = 100$ 和 200. 从下面模型中产生模拟数据

$$\begin{cases} Y_i = X_i^{\mathrm{T}}\beta_0 + g(T_i) + \sigma\varepsilon_i, \\ \widetilde{Y}_i = \phi(U_i)Y_i, \\ \widetilde{X}_{ir} = \psi_r(U_i)X_{ir}, \quad i = 1, \cdots, n, \quad r = 1, \cdots, 8, \end{cases}$$

其中参数向量真值取为 $\beta_0 = (2, 0, 1.5, 0, 0, 0, 1, 0)^{\mathrm{T}}$, $g(t) = 8t(1-t^2)$, 协变量 T 从区间 [0,1] 的均匀分布进行抽样, 失真变量 U 从区间 [5,7] 的均匀分布中产生. 预测变量和失真函数的产生为

$$X_1 \sim N(2, 1.144), \quad \psi_1(U) = \frac{U+10}{16}; \quad X_2 \sim N(0.5, 0.25), \quad \psi_2(U) = \frac{(U+1)^2}{50};$$

$$X_3 \sim N(1, 1), \quad \psi_3(U) = \frac{U+3}{9}; \qquad X_4 \sim N(0.8, 0.5), \quad \psi_4(U) = \frac{(U+2)^2}{65};$$

$$X_5 \sim N(1.2, 1.5), \quad \psi_5(U) = \frac{U+1}{7}; \quad X_6 \sim N(2, 1.5), \quad \psi_6(U) = \frac{(U-3)}{3};$$

$$X_7 \sim N(2, 2), \quad \psi_7(U) = \frac{U+5}{11}; \qquad X_8 \sim N(3, 1), \qquad \psi_8(U) = \frac{U-6}{2};$$

$$\varepsilon \sim N(0, 1), \qquad \phi(U) = \frac{(U+3)^2}{82}.$$

所有上面的失真函数都满足可识别性条件. 在模拟研究和实际数据分析中, 使用核函数 $K(t) = \frac{15}{32}(3 - 7t^2)I(|t| \leqslant 1)$ 来估计失真函数和非参数分量 $g(t)$. 类似于 Wang 和 Sun (2007), 窗宽 h 和 h_1 取为 $3\mathrm{sd}(U)n^{-1/3}\log(n)$ 和 $3\mathrm{sd}(U)n^{-1/5}$, 可以验证分别满足条件 (C5.5.2) 和条件 (C5.5.4), 其中 "sd" 表示样本的标准差. 如此取法而没有采用窗宽选择方法, 主要是为了节省计算时间. 对于给定的窗宽 h 和 h_1, 可以使用 Wang 等 (2009) 提出的 BIC 准则选取截断参数. 在模拟研究和实际数据分析中, 为了简单, 对于 SCAD 和 Lasso 等惩罚函数中所有的参数分量使用相同的截断参数 λ, 即 $\lambda_j = \lambda (j = 1, \cdots, p)$. 使用 Lasso 惩罚函数时, 截断参数取为 $\lambda_j = \lambda/|\hat{\beta}_j| \ \ (j = 1, \cdots, p)$, 其中 $\hat{\beta}_j \ \ (j = 1, \cdots, p)$ 是没有施加惩罚函数而得到的最小二乘估计. 因此, 可以极小化如下的 BIC 准则获得最优的截断参数 λ

$$\mathrm{BIC}(\lambda) = \log(\mathrm{MSE}(\lambda)) + \mathrm{DF}_\lambda(\log(n)/n),$$

其中 $\mathrm{MSE}(\lambda) = \dfrac{1}{n}\sum_{i=1}^{n}\left(\check{Y}_i - \check{X}_i^{\mathrm{T}}\hat{\beta}_\lambda\right)^2$, DF_λ 是 $\hat{\beta}_\lambda$ 中非零系数的个数, $\hat{\beta}_\lambda$ 表示具有截断参数 λ 的惩罚最小二乘目标函数的极小值, 定义于式 (5.5.7).

为了比较不同信噪比情况下的完成情况, 考虑三种不同的取值 $\sigma = 0.1$, 0.5 和 1. $\hat{g}(\cdot)$ 的表现情况用平均的均方根误差 (RASE) 来进行评价, 定义为

$$\mathrm{RASE} = \left\{\frac{1}{M}\sum_{l=1}^{M}\left[\hat{g}(t_l) - g(t_l)\right]^2\right\}^{1/2},$$

其中 $t_l, l = 1, \cdots, M$, 表示函数 $\hat{g}(t)$ 支撑集的格子点. 模拟中, 取 $M = 200$. 为了评价变量选择方法的完成情况, 比较了自适应的 Lasso 方法 (ALasso)、SCAD 方法、Lasso 方法和 oracle 估计. 模拟结果主要报告了零系数被正确估计成零的平均个数 (表中用 "C" 表示), 非零系数被错误估计成零的平均个数 (表中用 "IC" 表示). 类似于 Li 和 Liang (2008), 估计 $\hat{\beta}$ 的完成用广义均方误差 (GMSE) 进行评价, 定义为

$$\mathrm{GMSE} = E[(\hat{\beta} - \beta_0)^T E(XX^{\mathrm{T}})(\hat{\beta} - \beta_0)].$$

基于 1000 次重复模拟试验的结果报告在了表 5.5.1 中.

表 5.5.1　Lasso, SCAD 和自适应 Lasso (ALasso) 方法的模拟结果

方法	IC	C	GMSE	RASE	IC	C	GMSE	RASE
		($n = 100$, $\sigma = 0.1$)				($n = 200$, $\sigma = 0.1$)		
oracle	0	5.000	0.018	0.033	0	5.000	0.006	0.0075
Lasso	0	3.653	0.038	0.055	0	3.826	0.013	0.0106
SCAD	0	4.862	0.021	0.035	0	4.932	0.005	0.0079
ALasso	0	4.933	0.021	0.037	0	4.990	0.005	0.0077
		($n = 100$, $\sigma = 0.5$)				($n = 200$, $\sigma = 0.5$)		
oracle	0	5.000	0.035	0.067	0	5.000	0.010	0.0194
Lasso	0	3.478	0.087	0.120	0	3.648	0.031	0.0321
SCAD	0	4.833	0.056	0.074	0	4.915	0.014	0.0209
ALasso	0	4.855	0.062	0.078	0	4.960	0.015	0.0201
		($n = 100$, $\sigma = 1$)				($n = 200$, $\sigma = 1$)		
oracle	0	5.000	0.080	0.170	0	5.000	0.025	0.0581
Lasso	0.002	3.425	0.226	0.310	0	3.719	0.079	0.1034
SCAD	0	4.794	0.154	0.192	0	4.917	0.042	0.0613
ALasso	0	4.780	0.170	0.207	0	4.919	0.047	0.0623

表 5.5.1 中的结果说明了如下的情况.

(1) 随着样本量的增加或者噪声水平的减小, SCAD 和 Lasso 方法的估计误差

快速逼近到 oracle 估计的估计误差. 因此, 根据 C, IC, GMSE 和 RASE 的结果, 自适应的 Lasso 方法和 SCAD 方法完成结果明显好于 Lasso 方法.

(2) 从参数估计的 GMSE, 能够发现参数的估计几乎不影响非参数的估计. 该结果一致于理论结果, 回归参数的估计是 $\sqrt{n}$ 相合的, 函数分量的非参数估计是 $n^{2/5}$ 相合的.

(3) 噪声水平影响着估计结果, 并随着噪声水平的变大, GMSE 和 RASE 都变大.

一般而言, 根据零系数的平均个数、平均估计误差、GMSE 和 RASE、自适应 Lasso 方法和 SCAD 方法完成要优于 Lasso 方法.

2. 实际数据分析

本小节用肌酐数据 (creatinine data) 分析来说明所提的方法. 肌酐数据集包含年龄为 34~64 岁 188 个参与者在不同时间的重复测量数据, 详细试验设计的讨论见 Heaney 等 (1989). 一些研究者也对该数据使用协变量调整模型进行了分析, 如 Sentürk (2006), 与 Li 等 (2010a) 考虑该数据集来分析钙吸收 (absorbtion) 的均值建模.

钙吸收和钙的摄入量之间的关系是钙缺乏研究的主要兴趣问题. 初始研究表明年龄对有效的钙吸收有一个显著的非线性影响, 其他变量, 如体重指数 (body mass index, BMI)、膳食钙 (Cald)、尿钙 (Calu)、内源性粪 (Efc) 和磷 (Pho) 对响应变量钙吸收有线性影响. Heaney 等 (1989) 与 Heaney (2003) 研究发现体表面积 (BSA) 分别对钙的有效吸收和钙的摄入量有显著的影响. Sentürk (2006) 与 Li 等 (2010a) 建议把变量 BSA 作为混淆变量 U. 为了决定哪个变量对钙的有效吸收有显著性影响, 考虑如下模型:

$$\text{absorption} = \text{BMI}\beta_1 + \text{Cald}\beta_2 + \text{Calu}\beta_3 + \text{Efc}\beta_4 + \text{Pho}\beta_5 + g(\text{age}) + \varepsilon. \quad (5.5.8)$$

首先, 基于变量 absorption, BMI, Cald, Calu, Efc, Pho 和混淆变量 U, 所提估计程序可用于估计未知的失真函数 $\phi(U)$ 和 $\psi_r(U)$, $r = 1, \cdots, 5$. 其次, 使用惩罚最小二乘方法估计参数向量和未知函数 $g(\text{age})$. 基于 SCAD 惩罚函数和所提方法, 图 5.5.1 给了未知函数 $g(\text{age})$ 的拟合曲线 (由于 Lasso 惩罚函数的结果类似, 此处省略). 图 5.5.1 发现, 年龄对钙吸收有一个很强的非线性效应 (为一个 U 型曲线), 并发现钙吸收 35~55 岁有快速减少的趋势, 而在 55 岁以后有增长的趋势. 在表 5.5.2 中, 给出了普通最小二乘估计, Lasso 和 SCAD 惩罚函数的惩罚最小二乘估计的结果. 表 5.5.2 中的结果表明, 对于 Lasso 方法, 内源性粪 (Efc) 对钙吸收没有显著性影响, 而对于 SCAD 方法, 内源性粪 (Efc) 对钙吸收有负的影响. Lasso 和 SCAD 两种方法均说明 BMI 和 Pho 对钙吸收没有显著性影响, 而 Cald 和 Calu 分别对钙吸收有负的和正的影响.

表 5.5.2　肌酐数据 (creatinine data) 的估计和计算结果

变量	LS	Lasso	SCAD
BMI	0.0002	0	0
Cald	−0.1933	−0.1802	−0.1789
Calu	0.5334	0.4407	0.5369
Efc	−0.2527	0	−0.2180
Pho	0.0206	0	0

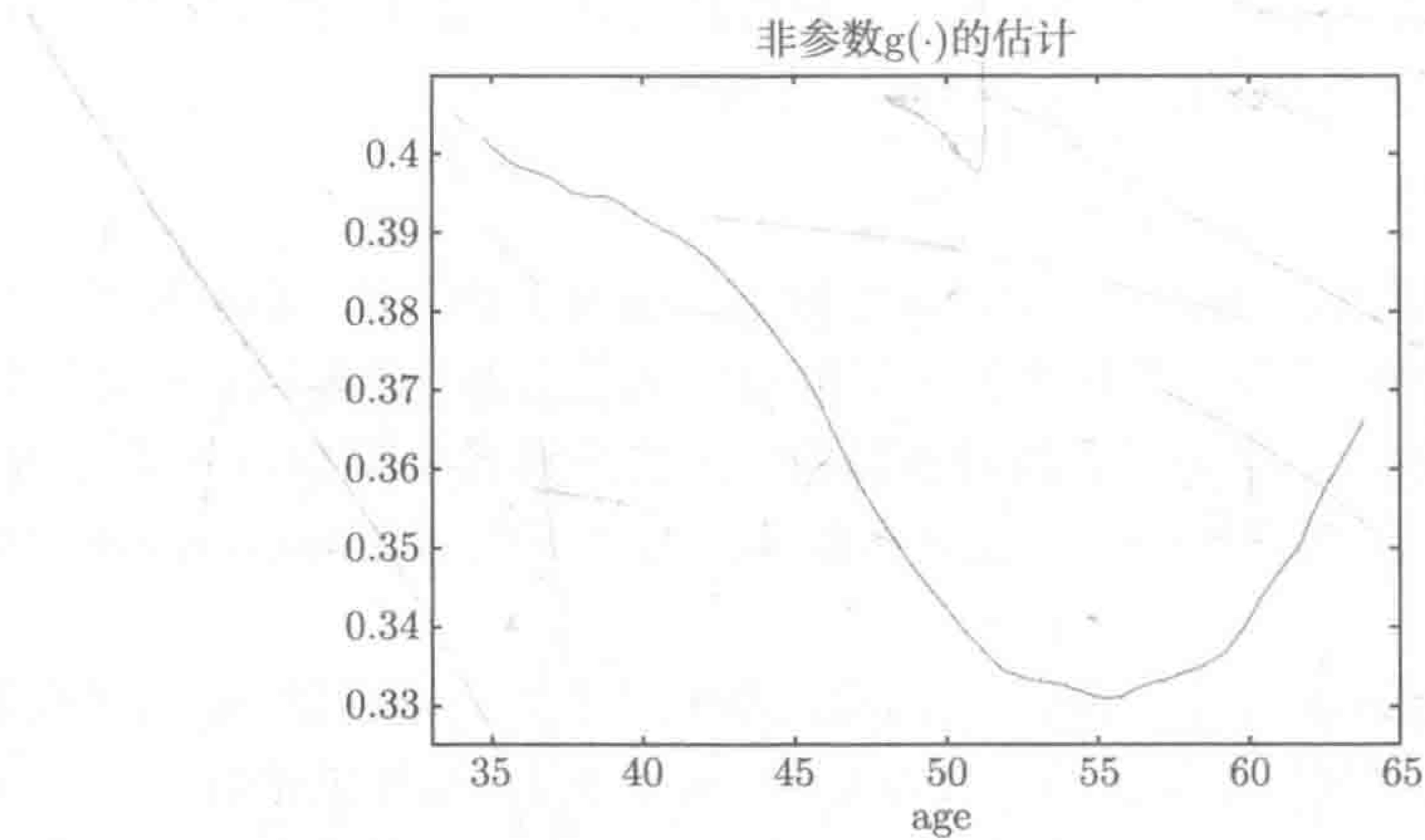


图 5.5.1　函数 $g(\text{age})$ 的拟合曲线

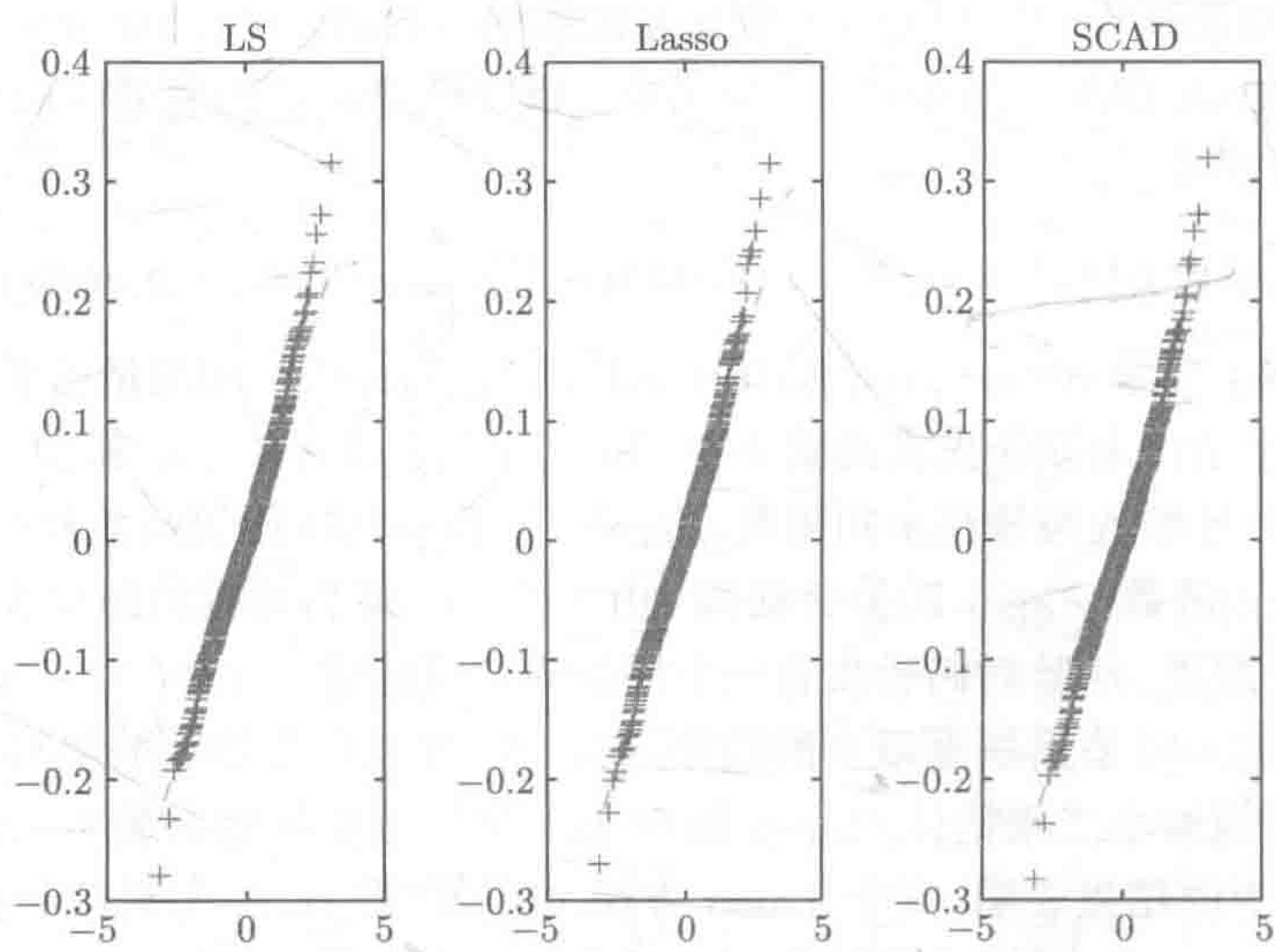


图 5.5.2　从左到右表示普通最小二乘方法、Lasso 方法和 SCAD 方法残差的正态 Q-Q 图

图 5.5.2 中残差的正态 Q-Q 图比较了普通最小二乘 (LS) 方法、Lasso 方法和

SCAD 方法, 进一步说明了使用模型 (5.5.8) 对该实际数据拟合的效果. 图 5.5.2 容易看到, 对照最小二乘方法的残差的正态 Q-Q 图, 即使变量 BMI 和 Pho 被 Lasso 和 SCAD 方法从最终的模型中排除, 但是 Lasso 和 SCAD 方法得到的残差是接近于普通最小二乘方法的残差, 并且所有的残差都渐近到正态分布. 因此, 残差的正态 Q-Q 图表明重要的变量被包含在最终的模型中, 说明所提方法完成得比较好.

5.6 协变量调整部分线性模型的约束估计和检验

5.6.1 改进的约束估计和渐近性质

5.5 节考虑了部分线性协变量调整回归模型 (5.5.1), 其中响应变量 Y 和协变量 X 全部受到混淆变量 U 的扭曲. 而在一些实际问题中, 遇到的只是 X 中的部分变量受到混淆变量 U 的扭曲或污染. 因此, 本节考虑如下一般的协变量调整部分线性模型:

$$\begin{cases} Y = X^{\mathrm{T}}\beta_0 + Z^{\mathrm{T}}\gamma_0 + g(T) + \varepsilon, \\ \widetilde{Y} = \phi(U)Y, \quad \widetilde{X} = \psi(U)X, \\ U\text{与}(X, Z, T, \varepsilon)\text{相互独立}, \end{cases} \tag{5.6.1}$$

其中 X 是协变量 $S = (X^{\mathrm{T}}, Z^{\mathrm{T}})^{\mathrm{T}}$ 中的 d 维协变量, $X \in \mathbb{R}^d$, $Z \in \mathbb{R}^q$, 且 $d + q = p$. 模型 (5.6.1) 中, $\widetilde{Y}, \widetilde{X}, Z, T, U$ 是可观测的变量, $\phi(\cdot)$ 和 $\psi(\cdot) = (\psi_1(\cdot), \cdots, \psi_d(\cdot))^{\mathrm{T}}$ 为未知的失真函数或扭曲函数. 进一步假定 $U \in \mathbb{R}$ 独立于变量 $(X^{\mathrm{T}}, Z^{\mathrm{T}}, Y, T, \varepsilon)$. 明显, 模型 (5.6.1) 是模型 (5.5.1) 的推广, 模型 (5.6.1) 表示协变量 S 中有一部分协变量 X 被混淆变量 U 的失真函数扭曲, 而剩余变量 Z 是被准确观测的. 当 $Z = 0$ 和 $d = p$ 时, 模型 (5.6.1) 就退化为模型 (5.5.1).

本节主要是研究模型 (5.6.1) 的估计和检验问题.

对模型 (5.6.1) 中的第一个模型, 两边关于 T 取条件数学期望, 并相减可以得到

$$Y - E(Y|T) = [X - E(X|T)]^{\mathrm{T}}\beta_0 + [Z - E(Z|T)]^{\mathrm{T}}\gamma_0 + \varepsilon. \tag{5.6.2}$$

模型 (5.6.2) 转变成了传统意义上的线性模型. 在总体均方误差意义下, 有

$$(\beta_0^{\mathrm{T}}, \gamma_0^{\mathrm{T}})^{\mathrm{T}} = \arg\min_{a,b} E\left\{Y - E(Y|T) - [X - E(X|T)]^{\mathrm{T}}a - [Z - E(Z|T)]^{\mathrm{T}}b\right\}^2,$$

可解得

$$\begin{pmatrix} \beta_0 \\ \gamma_0 \end{pmatrix} = E\left[\begin{pmatrix} X - E(X|T) \\ Z - E(X|T) \end{pmatrix}\begin{pmatrix} X - E(X|T) \\ Z - E(X|Z) \end{pmatrix}^{\mathrm{T}}\right]^{-1}$$
$$\times E\left[\begin{pmatrix} X - E(X|T) \\ Z - E(X|T) \end{pmatrix}(Y - E(Y|T))\right]. \tag{5.6.3}$$

如果变量 (X,Y) 能够被准确观测, 这时可利用核估计得到非参数的估计, 然后利用矩估计方法得到参数向量 $(\beta_0^{\mathrm{T}},\gamma_0^{\mathrm{T}})^{\mathrm{T}}$ 的估计量. 但在该模型中, 这两个变量不能观测到, 如果把相应的观测变量 $(\widetilde{X},\widetilde{Y})$ 直接替换变量 (X,Y) 时, 类似于前面几节的讨论可知, 得到的参数估计是不相合的.

下面介绍本节的估计方法. 假设 $\{(\widetilde{Y}_i,\widetilde{X}_i,Z_i,T_i,U_i), i=1,\cdots,n\}$ 是从模型 (5.6.1) 抽取的一组 i.i.d.的可观测样本. 要想有效地完成式 (5.6.3) 给出的参数向量 $(\beta_0^{\mathrm{T}},\gamma_0^{\mathrm{T}})^{\mathrm{T}}$ 的估计量, 首先需要估计不可观测的变量 (X,Y), 以及未知的函数 $g_Y(t)=E(Y|T=t), g_X(t)=E(X|T=t)$ 和 $g_Z(t)=E(Z|T=t)$. 关于不可观测变量 (X,Y) 估计, 在失真函数的约束条件下, 直接利用 5.5 节的直接估计方法可以得到. 利用 i.i.d. 的随机样本 $\{(\widetilde{Y}_i,\widetilde{X}_i,U_i), i=1,\cdots,n\}$, 即可用式 (5.5.2) 和式 (5.5.3) 得到 (X_i,Y_i) 的估计, 记为 $(\hat{X}_i,\hat{Y}_i), i=1,\cdots,n$. 首先回顾一下 Li 等 (2010a) 与 Li 等 (2014a) 对非参数函数 $g_Y(t)$ 和 $g_X(t)$ 的估计, 基于样本 $\{(\hat{Y}_i,\hat{X}_i,T_i), i=1,\cdots,n\}$, Li 等 (2010a) 与 Li 等 (2014a) 采用核估计方法对非参数函数 $g_Y(t)$ 和 $g_X(t)$ 的估计如下:

$$\hat{g}_Y(t)=\frac{\sum_{i=1}^n K_h(T_i-t)\hat{Y}_i}{\sum_{i=1}^n K_h(T_i-t)},\qquad \hat{g}_X(t)=\frac{\sum_{i=1}^n K_h(T_i-t)\hat{X}_i}{\sum_{i=1}^n K_h(T_i-t)}, \tag{5.6.4}$$

其中 $K_h(\cdot)=K(\cdot/h)/h$, h 是一个窗宽. 本节提出了一种新的估计非参数函数 $g_Y(t)$ 和 $g_X(t)$ 方法, 该方法有效提高了非参数函数的估计问题. 注意到根据 U 与 (X,Z,T,ε) 相互独立, 以及模型的识别性条件 $E[\phi(U)]=1, E[\psi_r(U)]=1$, 可直接建立如下等式

$$E(\widetilde{Y}|T)=E[\phi(U)Y|T]=E[\phi(U)]E(Y|T)=E(Y|T), \tag{5.6.5}$$
$$E(\widetilde{X}|T)=E[\psi(U)X|T]=E[\psi(U)]E(X|T)=E(X|T). \tag{5.6.6}$$

明显, 基于式 (5.6.5) 和式 (5.6.6), 可直接使用观测到的样本 $\{(\widetilde{Y}_i,\widetilde{X}_i,T_i), i=1,\cdots,n\}$ 来给出 $g_Y(t)=E(Y|T=t)$ 和 $g_X(t)=E(X|T=t)$ 的估计, 而不是使用估计的样本 $\{(\hat{Y}_i,\hat{X}_i,T_i), i=1,\cdots,n\}$ 对它们进行估计. 基于观测样本 $\{(\widetilde{Y}_i,\widetilde{X}_i,Z_i,T_i), i=$

$1,\cdots,n\}$ 和局部线性估计方法, 非参数 $g_Y(t)$, $g_X(t)$ 和 $g_Z(t)$ 的估计分别如下:

$$\begin{cases} \hat{g}_{\widetilde{Y}}(t)=:\hat{E}(\widetilde{Y}|T=t)=\dfrac{S_{n2}(t)Q_{n0,\widetilde{Y}}(t)-S_{n1}(t)Q_{n1,\widetilde{Y}}(t)}{S_{n2}(t)S_{n0}(t)-S_{n1}^2(t)}, \\ \hat{g}_{\widetilde{X}}(t)=:\hat{E}(\widetilde{X}|T=t)=\dfrac{S_{n2}(t)Q_{n0,\widetilde{X}}(t)-S_{n1}(t)Q_{n1,\widetilde{X}}(t)}{S_{n2}(v)S_{n0}(t)-S_{n1}^2(t)}, \\ \hat{g}_{Z}(t)=:\hat{E}(Z|T=t)=\dfrac{S_{n2}(t)Q_{n0,Z}(t)-S_{n1}(t)Q_{n1,Z}(t)}{S_{n2}(t)S_{n0}(t)-S_{n1}^2(t)}. \end{cases} \tag{5.6.7}$$

其中

$$S_{n\omega}(t)=\frac{1}{nh}\sum_{i=1}^{n}\left(\frac{T_i-t}{h}\right)^{\omega}K\left(\frac{T_i-t}{h}\right),$$

$$Q_{n\omega,L}(t)=\frac{1}{nh}\sum_{i=1}^{n}\left(\frac{T_i-t}{h}\right)^{\omega}K\left(\frac{T_i-t}{h}\right)L_i,\ \omega=0,1,2.$$

这时, 由式 (5.6.3), 参数向量 $(\beta_0^{\mathrm{T}},\gamma_0^{\mathrm{T}})^{\mathrm{T}}$ 在样本情形下最小二乘估计为

$$\begin{aligned}\begin{pmatrix}\hat{\beta}\\ \hat{\gamma}\end{pmatrix}=&\left[\frac{1}{n}\sum_{i=1}^{n}\begin{pmatrix}\hat{X}_i-\hat{g}_{\widetilde{X}}(T_i)\\ Z_i-\hat{g}_Z(T_i)\end{pmatrix}\begin{pmatrix}\hat{X}_i-\hat{g}_{\widetilde{X}}(T_i)\\ Z_i-\hat{g}_Z(T_i)\end{pmatrix}^{\mathrm{T}}\right]^{-1}\\ &\times\left[\frac{1}{n}\sum_{i=1}^{n}\begin{pmatrix}\hat{X}_i-\hat{g}_{\widetilde{X}}(T_i)\\ Z_i-\hat{g}_Z(T_i)\end{pmatrix}\left[\hat{Y}_i-\hat{g}_{\widetilde{Y}}(T_i)\right]\right].\end{aligned} \tag{5.6.8}$$

式 (5.6.8) 给出的最小二乘估计是在识别性约束条件 $E[\phi(U)]=1$ 和 $E[\psi_r(U)]=1$ 下得到的估计量, 因此把该估计称为改进的约束估计. 明显, 该估计量不同于 Li 等 (2010a) 所定义的直接估计.

注意到 $E[\widetilde{Y}-\widetilde{X}^{\mathrm{T}}\beta_0-Z^{\mathrm{T}}\gamma_0|T=t]=g(t)$, 则再次使用观测样本 $\{(\widetilde{Y}_i,\widetilde{X}_i,Z_i,T_i),$ $i=1,\cdots,n\}$, 式 (5.6.8) 给出的参数向量的最小二乘估计 $(\hat{\beta}^{\mathrm{T}},\hat{\gamma}^{\mathrm{T}})^{\mathrm{T}}$, 利用局部线性估计方法, 可得 $g(t)$ 的估计如下

$$\hat{g}(t,\hat{\beta},\hat{\gamma})=\frac{S_{n2}(t)D_{n0}(t,\hat{\beta},\hat{\gamma})-S_{n1}(t)D_{n1}(t,\hat{\beta},\hat{\gamma})}{S_{n2}(t)S_{n0}(t)-S_{n1}^2(t)}, \tag{5.6.9}$$

其中 $D_{n\omega}(t,\hat{\beta},\hat{\gamma})=\dfrac{1}{nh}\sum_{i=1}^{n}K\left(\dfrac{T_i-t}{h}\right)\left(\dfrac{T_i-t}{h}\right)^{\omega}\left[\widetilde{Y}_i-\widetilde{X}_i^{\mathrm{T}}\hat{\beta}-Z_i^{\mathrm{T}}\hat{\gamma}\right]$, $\omega=0,1$. 为了得到 $\hat{\beta},\hat{\gamma},\hat{g}(t,\hat{\beta},\hat{\gamma})$ 的渐近结果, 需要如下的假设条件.

(C5.6.1) 模型误差 ε 三阶矩有限: $E[|\varepsilon|^3]<+\infty$.

(C5.6.2) 密度函数 $f_U(u),u\in\mathcal{U},f_T(t),t\in\mathcal{T}$ 大于零, 并且都满足一阶 Lipschitz 条件. 这里的 $\mathcal{U}$ 和 $\mathcal{T}$ 分别表示随机变量 U 和 T 的紧支撑. 此外, 还需要求 $\inf\limits_{u\in\mathcal{U}}f_U(u)\geqslant c_0$, 其中 $0<c_0<+\infty$.

(C5.6.3) 函数 $\phi(u), \psi_r(u), g_Y(t), g_{X,r}(t)$ 和 $g_{Z,s}(t)$ 有连续有界的三阶导数, 其中 $g_{X,r}(t)$ 表示 $g_X(t)$ 的第 r 个分量, $g_{Z,s}(t)$ 表示 $g_Z(t)$ 的第 s 个分量, 且 $r=1,\cdots,d$ 和 $s=1,\cdots,q$. 此外, $\phi(u)$ 与 $\psi_r(u)$ 在 $u\in\mathcal{U}$ 上取值不为零.

(C5.6.4) $E[Y]$ 和 $E[X_r]$ 都不等于零, 且 $E[|Y|^3]<\infty$, $E[|X_r|^3]<\infty$, $r=1,\cdots,d$.

(C5.6.5) 核函数 $K(\cdot)$是定义在紧支撑集上的一个关于原点对称的密度函数, $K(\cdot)$ 有界并满足 Lipschitz 条件, $\mu_{K_2}=\int_{-\infty}^{\infty}u^2K(u)\mathrm{d}u\neq 0$, $\int_{-\infty}^{\infty}|u|^jK(u)\mathrm{d}u<\infty$, $j=1,2,3$.

(C5.6.6) 当样本量大小 $n\to\infty$ 时, 窗宽 h, h_* 满足:

(1) $\dfrac{\log^2 n}{nh_*^2}\to 0$, $nh_*^2\to\infty$, $nh_*^4\to 0$.

(2) $\dfrac{\log^2 n}{nh^2}\to 0$, $nh^2\to\infty$, $nh^8\to 0$.

上面条件中, 条件 (C5.6.1)~条件 (C5.6.5) 都是半参数模型的研究中常用的条件, 此处仅对条件 (C5.6.6) 进行说明. 在条件 (C5.6.6) 中, 需要分别对窗宽 h_* 和 h 进行选取. 注意到条件 $nh_*^4\to 0$ 表明在得到估计 $\hat{X}_i$ 和 $\hat{Y}_i$ 时需要对窗宽 h_* 需要进行 "欠光滑"(undersmoothing). 因此, Wei 等 (2015) 建议使用由 Silverman (1986), Zhou 和 Liang (2008), Zhang 等 (2014) 等使用的拇指法则 (rule of thumb) 进行选取窗宽, 即 $h_*=\hat{\sigma}_U n^{-1/3}$, 其中 $\hat{\sigma}_U$ 是混淆变量 U 的样本标准差. 对于窗宽 h 的选取, 条件 (C5.6.6) (2) 表明在估计 $g_Y(t)$, $g_X(t)$ 和 $g_Z(t)$ 时, h 可以取到最优窗宽 $n^{-1/5}$. 因此, 可以使用交错核实 (cross-validation) 的方法来选取 h. 下面就以估计 $g_Y(t)=E(Y|T=t)$ 为例进行说明.

$$\mathrm{CV}_{\widetilde{Y}}(h_0)=n^{-1}\sum_{i=1}^{n}\left\{\widetilde{Y}_i-\hat{g}_{\widetilde{Y},h_0}^{(-i)}(T_i)\right\}^2,$$

其中 $\hat{g}_{\widetilde{Y},h_0}^{(-i)}(t)$ 是将数据 $\{\widetilde{Y}_i,T_i\}_{i=1}^n$ 中的第 i 个样本观测值 $(\widetilde{Y}_i,T_i)$ 去掉之后, 使用事先给定的窗宽 h_0, 再使用与式 (5.6.8) 中相同的局部线性估计方法得到. 此时, h 最终可选取为 $h_Y=\arg\min\limits_{h_0}\mathrm{CV}_{\widetilde{Y}}(h_0)$. 此外, 对于估计 $g_X(t)$ 和 $g_Z(t)$ 中涉及的窗宽选择, 可类似地提出. 接下来, 定理 5.6.1 给出估计 $\hat{\beta},\hat{\gamma}$ 的渐近结果.

定理 5.6.1　假设正则条件(C5.6.1)~条件(C5.6.6)成立, 当 $n\to\infty$ 时, 则有

$$\begin{aligned}\begin{pmatrix}\hat{\beta}\\ \hat{\gamma}\end{pmatrix}=&\begin{pmatrix}\beta_0\\ \gamma_0\end{pmatrix}+\frac{1}{n}\sum_{i=1}^{n}\Gamma_{\check{S}}^{-1}\check{S}_i\varepsilon_i+\frac{1}{n}\sum_{i=1}^{n}\Gamma_{\check{S}}^{-1}\left(\frac{\widetilde{Y}_i-Y_i}{E[Y]}\right)E\left[\check{S}Y\right]\\&+\frac{1}{n}\sum_{i=1}^{n}\sum_{r=1}^{d}\Gamma_{\check{S}}^{-1}\left(\frac{\widetilde{X}_{ri}-X_{ri}}{E[X_r]}\right)\beta_{0,r}E\left[\check{S}X_r\right]+o_P(n^{-1/2}),\end{aligned}$$

其中 $S=(X^{\mathrm{T}},Z^{\mathrm{T}})^{\mathrm{T}}, \breve{S}=S-E[S|T], \breve{S}_i=S_i-E[S_i|T_i], i=1,\cdots,n, \Gamma_{\breve{S}}=E(\breve{S}\breve{S}^{\mathrm{T}})$, $\beta_{0,r}$ 为参数 β_0 的第 r 个分量, $r=1,\cdots,d$.

对非参数函数估计 $\hat{g}(t,\hat{\beta},\hat{\gamma})$, 有如下渐近表达式. 定理 5.6.2 同时表明, 若失真函数 $\phi(\cdot)\neq 1, \psi_r(\cdot)\neq 1$, 此时混淆变量 U 对 $g(t)$ 的估计也有影响, 增大了估计的渐近方差.

定理 5.6.2 假设正则条件(C5.6.1)~条件(C5.6.6)成立, 当 $n\to\infty$ 时, 则对 $t\in\mathcal{T}$ 一致地有

$$\hat{g}(t,\hat{\beta},\hat{\gamma})=g(t)+\frac{1}{nhf_T(t)}\sum_{i=1}^{n}K\left(\frac{T_i-t}{h}\right)\left[\varepsilon_i+(\widetilde{Y}_i-Y_i)+(X_i-\widetilde{X}_i)^{\mathrm{T}}\beta_0\right]$$
$$+\frac{1}{2}g''(t)h^2\mu_{K_2}+o_P\left(h^2+\sqrt{\frac{\log n}{nh}}\right).$$

5.6.2 假设检验

本节考虑该模型 (5.6.1) 下的假设检验问题. 为了方便, 定义 $\theta_0=(\beta_0^{\mathrm{T}},\gamma_0^{\mathrm{T}})^{\mathrm{T}}$.

$$H_0: A\theta_0=b \quad \text{vs} \quad H_1: A\theta_0\neq b, \tag{5.6.10}$$

其中 A 是一个 $k\times p$ 的已知矩阵, 其秩 $\mathrm{rank}\,(A)=k\leqslant p$, b 是一个 k 维的已知常数向量. 在很多实际问题的研究中, 有时需要对模型的参数向量 (或者变量之间存在着一些线性关系) 附加一定的约束条件, 带有约束问题的参数估计与假设检验在计量经济学中尤为常见. 如果能在统计分析的过程中充分利用该约束条件, 则有可能得到更加有效的估计, 从而提高估计的精度.

原假设为 H_0 为真的情况下, 原假设中包含的参数信息可作为真实的约束条件. 一个自然而然的想法就是使用带约束条件下的最大最小化估计问题. 因此, 采用拉格朗日算子构造在线性约束条件 $A\theta_0=b$ 下的限制约束估计 $\hat{\theta}_R=(\hat{\beta}_R^{\mathrm{T}},\hat{\gamma}_R^{\mathrm{T}})^{\mathrm{T}}$, 即关于 θ 极小化

$$\mathcal{L}(\theta,\lambda)=\sum_{i=1}^{n}\left[\hat{Y}_i-\hat{g}_{\widetilde{Y}}(T_i)-\begin{pmatrix}\hat{X}_i-\hat{g}_{\widetilde{X}}(T_i)\\ Z_i-\hat{g}_Z(T_i)\end{pmatrix}^{\mathrm{T}}\theta\right]^2+2\lambda^{\mathrm{T}}(A\theta-b).$$

关于 $\mathcal{L}(\theta,\lambda)$ 求偏导数, 可以得到

$$\begin{cases}\dfrac{\partial\mathcal{L}(\theta,\lambda)}{\partial\theta}=-2\displaystyle\sum_{i=1}^{n}\begin{pmatrix}\hat{X}_i-\hat{g}_{\widetilde{X}}(T_i)\\ Z_i-\hat{g}_Z(T_i)\end{pmatrix}\left[\hat{Y}_i-\hat{g}_{\widetilde{Y}}(T_i)-\begin{pmatrix}\hat{X}_i-\hat{g}_{\widetilde{X}}(T_i)\\ Z_i-\hat{g}_Z(T_i)\end{pmatrix}^{\mathrm{T}}\theta\right]\\ \qquad +2A^{\mathrm{T}}\lambda=0,\\ \dfrac{\partial\mathcal{L}(\theta,\lambda)}{\partial\lambda}=2(A\theta-b)=0.\end{cases}$$

求解上面方程, 可以得到

$$\begin{aligned}\begin{pmatrix}\hat{\beta}_R\\ \hat{\gamma}_R\end{pmatrix}=\begin{pmatrix}\hat{\beta}\\ \hat{\gamma}\end{pmatrix}-&\left[\sum_{i=1}^n\begin{pmatrix}\hat{X}_i-\hat{g}_{\tilde{X}}(T_i)\\ Z_i-\hat{g}_Z(T_i)\end{pmatrix}\begin{pmatrix}\hat{X}_i-\hat{g}_{\tilde{X}}(T_i)\\ Z_i-\hat{g}_Z(T_i)\end{pmatrix}^{\mathrm{T}}\right]^{-1}\\ &\times A^{\mathrm{T}}\left\{A\left[\sum_{i=1}^n\begin{pmatrix}\hat{X}_i-\hat{g}_{\tilde{X}}(T_i)\\ Z_i-\hat{g}_Z(T_i)\end{pmatrix}\begin{pmatrix}\hat{X}_i-\hat{g}_{\tilde{X}}(T_i)\\ Z_i-\hat{g}_Z(T_i)\end{pmatrix}^{\mathrm{T}}\right]^{-1}A^{\mathrm{T}}\right\}^{-1}\\ &\times\left[A\begin{pmatrix}\hat{\beta}\\ \hat{\gamma}\end{pmatrix}-b\right].\end{aligned}\tag{5.6.11}$$

式 (5.6.11) 中的 $(\hat{\beta}^{\mathrm{T}},\hat{\gamma}^{\mathrm{T}})^{\mathrm{T}}$ 是由式 (5.6.8) 给出. 对于限制约束估计 $(\hat{\beta}_R^{\mathrm{T}},\hat{\gamma}_R^{\mathrm{T}})^{\mathrm{T}}$, 有如下的渐近结果.

定理 5.6.3　假设正则条件(C5.6.1)~条件(C5.6.6)成立, 当 $n\to\infty$ 时, 则有

$$\sqrt{n}\begin{pmatrix}\hat{\beta}_R-\beta_0\\ \hat{\gamma}_R-\gamma_0\end{pmatrix}\stackrel{L}{\longrightarrow}N\left(0,\mathcal{O}_A\Gamma_{\breve{S}}^{-1}\Lambda_{\breve{S}}^{\varepsilon}\Gamma_{\breve{S}}^{-1}\mathcal{O}_A^{\mathrm{T}}+\mathcal{O}_AQ_{\beta_0}\mathcal{O}_A^{\mathrm{T}}\right),$$

其中 $\mathcal{O}_A=I_p-\Gamma_{\breve{S}}^{-1}A^{\mathrm{T}}\left(A\Gamma_{\breve{S}}^{-1}A^{\mathrm{T}}\right)^{-1}A$, I_p 是一个维数为 p 的单位矩阵, $\Lambda_{\breve{S}}^{\varepsilon}=E\left(\breve{S}\breve{S}^{\mathrm{T}}\varepsilon^2\right)$, 以及

$$\begin{aligned}Q_{\beta_0}=&\sum_{l=1}^d\sum_{r=1}^d\beta_{0,l}\beta_{0,r}\Gamma_{\breve{S}}^{-1}\Lambda_{\breve{S},l}\Lambda_{\breve{S},r}^{\mathrm{T}}\Gamma_{\breve{S}}^{-1}\mathrm{Cov}\left(\psi_l(U),\psi_r(U)\right)\frac{E(X_lX_r)}{E(X_l)E(X_r)}\\ &-\sum_{l=1}^d\beta_{0,l}\Gamma_{\breve{S}}^{-1}\left[\Lambda_{\breve{S},l}\Lambda_{\breve{S},Y}^{\mathrm{T}}+\Lambda_{\breve{S},Y}\Lambda_{\breve{S},l}^{\mathrm{T}}\right]\Gamma_{\breve{S}}^{-1}\mathrm{Cov}\left(\psi_l(U),\phi(U)\right)\frac{E(X_lY)}{E(X_l)E(Y)}\\ &+\Gamma_{\breve{S}}^{-1}\Lambda_{\breve{S},Y}\Lambda_{\breve{S},Y}^{\mathrm{T}}\Gamma_{\breve{S}}^{-1}\mathrm{Var}\left(\phi(U)\right)\frac{E\left(Y^2\right)}{[E(Y)]^2},\end{aligned}$$

其中 $\Lambda_{\breve{S},r}=E(\breve{S}X_r)$, $r=1,\cdots,d$, 以及 $\Lambda_{\breve{S},Y}=E(\breve{S}Y)$.

下面给出判断检验原假设 H_0 是否成立的的检验统计量. 该检验统计量的想法是基于原假设与全局条件下的残差平方和之差. 首先构造原假设下的残差平方和

$$\mathrm{RSS}(H_0)=\sum_{i=1}^n\left[\hat{Y}_i-\hat{X}_i^{\mathrm{T}}\hat{\beta}_R-Z_i^{\mathrm{T}}\hat{\gamma}_R-\hat{g}(T_i,\hat{\beta}_R,\hat{\gamma}_R)\right]^2.\tag{5.6.12}$$

再构造全局残差平方和

$$\mathrm{RSS}(H_1)=\sum_{i=1}^n\left[\hat{Y}_i-\hat{X}_i^{\mathrm{T}}\hat{\beta}-Z_i^{\mathrm{T}}\hat{\gamma}-\hat{g}(T_i,\hat{\beta},\hat{\gamma})\right]^2.\tag{5.6.13}$$

Wei 等 (2015) 提出的检验统计量为

$$\mathcal{T}_n = \mathrm{RSS}(H_0) - \mathrm{RSS}(H_1). \tag{5.6.14}$$

如果原假设 H_0 为真, 该检验统计量 $\mathcal{T}_n$ 倾向于取较小的值. 如果原假设 H_0 不为真, 该检验统计量 $\mathcal{T}_n$ 倾向于取较大的值. 那么该统计量 $\mathcal{T}_n$ 取值的相对大小, 就可以作为判断原假设是否为真的一个依据.

定理 5.6.4 假设正则条件(C5.6.1)~条件(C5.6.6)成立, 当 $n \to \infty$ 时, 则有下面的结论.

(1) 原假设 $H_0: A\theta_0 = b$ 为真时, 有

$$\mathcal{T}_n \xrightarrow{L} \ell_1\chi^2_{1,1} + \ell_2\chi^2_{2,1} + \cdots + \ell_k\chi^2_{k,1},$$

其中 $\chi^2_{i,1}, i = 1, \cdots, k$ 是独立同分布自由度为 1 的 χ^2 变量, ℓ_i 是矩阵

$$\left(A\left[\Gamma^{-1}_{\breve{S}}\Lambda^{\varepsilon}_{\breve{S}}\Gamma^{-1}_{\breve{S}}\sigma^2_{\varepsilon} + Q_{\beta_0}\right]A^{\mathrm{T}}\right)\left(A\Gamma^{-1}_{\breve{S}}A^{\mathrm{T}}\right)^{-1}$$

的特征值.

(2) 在局部备择假设 $H_{1n}: A\theta_0 = b + n^{-1/2}c$ 为真时, $\mathcal{T}_n$服从广义 χ^2 分布 (Sheil and O'Muircheartaigh, 1977), 即

$$\mathcal{T}_n \xrightarrow{L} \left(\mathcal{K} + \Gamma^{-1/2}A^{\mathrm{T}}\left[A\Gamma^{-1}_{\breve{S}}A^{\mathrm{T}}\right]^{-1}c\right)^{\mathrm{T}}\left(\mathcal{K} + \Gamma^{-1/2}_{\breve{S}}A^{\mathrm{T}}\left[A\Gamma^{-1}_{\breve{S}}A^{\mathrm{T}}\right]^{-1}c\right),$$

其中 $\mathcal{K}$ 服从多元正态分布 $N(0, H)$, 该分布的渐近方差 H 有如下的表达式

$$H = \Gamma^{-1/2}_{\breve{S}}A^{\mathrm{T}}\left(A\Gamma^{-1}_{\breve{S}}A^{\mathrm{T}}\right)^{-1}A\left(\Gamma^{-1}_{\breve{S}}\Lambda^{\varepsilon}_{\breve{S}}\Gamma^{-1}_{\breve{S}} + Q_{\beta_0}\right)A^{\mathrm{T}}\left(A\Gamma^{-1}_{\breve{S}}A^{\mathrm{T}}\right)^{-1}A\Gamma^{-1/2}_{\breve{S}}.$$

这里虽然给出了统计量 $\mathcal{T}_n$ 的渐近分布, 但分布中未知的参数过多, 使得在有限样本下的数值模拟表现结果不理想. 为此, 提出了 wild-bootstrap 方法. 具体过程如下.

步骤 1 计算统计量 $\mathcal{T}_n$.

步骤 2 随机产生 B 组独立同分布的二项分布随机变量 $\delta_{il}, i = 1, \cdots, n, l = 1, \cdots, B$, 其分布为 $P\left(\delta_{il} = \dfrac{1 \mp \sqrt{5}}{2}\right) = \dfrac{5 \pm \sqrt{5}}{10}$, 并计算残差 $\hat{\varepsilon}_i = \hat{Y}_i - \hat{X}_i^{\mathrm{T}}\hat{\beta} - Z_i^{\mathrm{T}}\hat{\gamma} - \hat{g}(T_i, \hat{\beta}, \hat{\gamma})$ 与 bootstrap 残差 $\hat{\varepsilon}_i^{(l)} = \hat{\varepsilon}_i\delta_{il}$, 再用这些估计量与残差得到新的响应变量 $\hat{Y}_i^{(l)}$:

$$\hat{Y}_i^{(l)} = \hat{X}_i^{\mathrm{T}}\hat{\beta} + Z_i^{\mathrm{T}}\hat{\gamma} + \hat{g}(T_i, \hat{\beta}, \hat{\gamma}) + \hat{\varepsilon}_i^{(l)}.$$

步骤 3 使用数据 $(\hat{Y}_i^{(l)}, \hat{X}_i, Z_i)$, 得到新的估计 $(\hat{\beta}^{(l)\mathrm{T}}, \hat{\gamma}^{(l)\mathrm{T}})^{\mathrm{T}}$, $(\hat{\beta}_R^{(l)\mathrm{T}}, \hat{\gamma}_R^{(l)\mathrm{T}})^{\mathrm{T}}$ 分别如下:

$$\begin{pmatrix} \hat{\beta}^{(l)} \\ \hat{\gamma}^{(l)} \end{pmatrix} = \left[\frac{1}{n} \sum_{i=1}^{n} \begin{pmatrix} \hat{X}_i - \hat{g}_{\widetilde{X}}(T_i) \\ Z_i - \hat{g}_Z(T_i) \end{pmatrix} \begin{pmatrix} \hat{X}_i - \hat{g}_{\widetilde{X}}(T_i) \\ Z_i - \hat{g}_Z(T_i) \end{pmatrix}^{\mathrm{T}} \right]^{-1}$$
$$\times \left[\frac{1}{n} \sum_{i=1}^{n} \begin{pmatrix} \hat{X}_i - \hat{g}_{\widetilde{X}}(T_i) \\ Z_i - \hat{g}_Z(T_i) \end{pmatrix} \left[\hat{Y}_i^{(l)} - \hat{g}_{\widetilde{Y}}(T_i) \right] \right],$$

$$\begin{pmatrix} \hat{\beta}_R^{(l)} \\ \hat{\gamma}_R^{(l)} \end{pmatrix} = \begin{pmatrix} \hat{\beta}^{(l)} \\ \hat{\gamma}^{(l)} \end{pmatrix} - \left[\sum_{i=1}^{n} \begin{pmatrix} \hat{X}_i - \hat{g}_{\widetilde{X}}(T_i) \\ Z_i - \hat{g}_Z(T_i) \end{pmatrix} \begin{pmatrix} \hat{X}_i - \hat{g}_{\widetilde{X}}(T_i) \\ Z_i - \hat{g}_Z(T_i) \end{pmatrix}^{\mathrm{T}} \right]^{-1} A^{\mathrm{T}}$$
$$\times \left\{ A \left[\sum_{i=1}^{n} \begin{pmatrix} \hat{X}_i - \hat{g}_{\widetilde{X}}(T_i) \\ Z_i - \hat{g}_Z(T_i) \end{pmatrix} \begin{pmatrix} \hat{X}_i - \hat{g}_{\widetilde{X}}(T_i) \\ Z_i - \hat{g}_Z(T_i) \end{pmatrix}^{\mathrm{T}} \right]^{-1} A^{\mathrm{T}} \right\}^{-1}$$
$$\times \left[A \begin{pmatrix} \hat{\beta}^{(l)} \\ \hat{\gamma}^{(l)} \end{pmatrix} - b \right].$$

进一步计算 bootstrap 统计量 $\mathcal{T}_n^{(l)}$

$$\mathcal{T}_n^{(l)} = \mathrm{RSS}^{(l)}(H_0) - \mathrm{RSS}^{(l)}(H_1),$$

其中

$$\mathrm{RSS}^{(l)}(H_0) = \sum_{i=1}^{n} \left[\hat{Y}_i^{(l)} - \hat{X}_i^{\mathrm{T}} \hat{\beta}_R^{(l)} - Z_i^{\mathrm{T}} \hat{\gamma}_R^{(l)} - \hat{g}(T_i, \hat{\beta}_R^{(l)}, \hat{\gamma}_R^{(l)}) \right]^2$$

和

$$\mathrm{RSS}^{(l)}(H_1) = \sum_{i=1}^{n} \left[\hat{Y}_i^{(l)} - \hat{X}_i^{\mathrm{T}} \hat{\beta}^{(l)} - Z_i^{\mathrm{T}} \hat{\gamma}^{(l)} - \hat{g}(T_i, \hat{\beta}^{(l)}, \hat{\gamma}^{(l)}) \right]^2.$$

步骤 4 最后计算 bootstrap 统计量 $\mathcal{T}_n^{(l)}$ 的 $1-\alpha$ 样本分位数, 并将该分位数记为 α 水平下的分位点.

5.6.3 数值模拟

本节通过数值模拟来考察所提到估计与检验方法. 核函数选取为 Epanechnikov 核函数: $K(t) = 0.75(1-t^2)I(|t| \leqslant 1)$. 窗宽 h_* 的选取按照提出的拇指法则方法进行选取, h 的选取采用交错核实方法.

模拟数据来自于如下的部分线性模型:

$$Y = \beta_{0,1}X_1 + \beta_{0,2}X_2 + \gamma_{0,1}Z_1 + 2\sin(2\pi T) + \varepsilon, \tag{5.6.15}$$

其中 $(\beta_{0,1}, \beta_{0,2}, \gamma_{0,1})^{\mathrm{T}} = (1, -1, 1.5)^{\mathrm{T}}$. 重复模拟 500 次, 样本容量分别取 $n = 300, 500, 1000$, $X_1 \sim N(2, 1.2^2)$, $X_2 \sim N(1, 1)$, $Z_1 \sim N(0.5, 0.25^2)$, T 来自区间

$(0,1)$ 上的均匀分布. 模型误差 ε 与变量 (X_1, X_2, Z_1, T) 相互独立, $\varepsilon \sim N(0, \sigma^2)$, 其中 $\sigma = 0.25, 0.75$. 混淆变量 U 来自于区间 $(0,1)$ 上的均匀分布, 并且与 $(X_1, X_2, Z_1, T, \varepsilon)$ 相互独立. 失真函数选取为 $\psi_1(U) = 1 + 0.3\cos(2\pi U)$, $\psi_2(U) = 1 + 0.2\left(U^2 - \frac{1}{3}\right)$, $\phi(U) = 1 + 0.2\cos(2\pi U)$.

模拟研究主要针对本节所提出的改进的约束估计方法和 Li 等 (2010a) 提出的直接估计方法进行了比较, 模拟结果见表 5.6.1, 表中提供了在 $\sigma = 0.25$ 和 $\sigma = 0.75$ 两种情况下, 参数估计的均值 (Mean)、标准差 (SD) 和均方误差 (MSE).

表 5.6.1 模型 (5.6.15) 中参数估计的均值 (Mean)、标准差 (SD) 和均方误差 (MSE)

σ	n	参数	约束估计			直接估计		
			Mean	SD	MSE	Mean	SD	MSE
$\sigma = 0.25$	300	$\beta_{0,1} = 1$	1.0112	0.0316	0.0012	1.0121	0.0320	0.0013
		$\beta_{0,2} = -1$	−0.9900	0.0415	0.0022	−0.9902	0.0415	0.0022
		$\gamma_{0,1} = 1.5$	1.5101	0.1110	0.0135	1.5100	0.1110	0.0135
	500	$\beta_{0,1} = 1$	1.0056	0.0178	0.0005	1.0062	0.0180	0.0005
		$\beta_{0,2} = -1$	−0.9937	0.0257	0.0007	−0.9937	0.0257	0.0007
		$\gamma_{0,1} = 1.5$	1.5088	0.0736	0.0051	1.5088	0.0735	0.0051
	1000	$\beta_{0,1} = 1$	1.0025	0.0110	0.0001	1.0028	0.0111	0.0001
		$\beta_{0,2} = -1$	−1.0004	0.0149	0.0003	−1.0005	0.0150	0.0003
		$\gamma_{0,1} = 1.5$	1.5048	0.0460	0.0019	1.5048	0.0460	0.0019
$\sigma = 0.75$	300	$\beta_{0,1} = 1$	1.0121	0.0478	0.0088	1.0131	0.0480	0.0092
		$\beta_{0,2} = -1$	−0.9942	0.0597	0.0043	−0.9947	0.0597	0.0043
		$\gamma_{0,1} = 1.5$	1.5157	0.2114	0.0922	1.5158	0.2116	0.0929
	500	$\beta_{0,1} = 1$	1.0076	0.0315	0.0011	1.0081	0.0317	0.0012
		$\beta_{0,2} = -1$	−0.9981	0.0440	0.0016	−0.9983	0.0440	0.0016
		$\gamma_{0,1} = 1.5$	1.5040	0.1497	0.0211	1.5040	0.1498	0.0211
	1000	$\beta_{0,1} = 1$	1.0014	0.0228	0.0005	1.0017	0.0229	0.0005
		$\beta_{0,2} = -1$	−1.0010	0.0280	0.0008	−1.0010	0.0280	0.0008
		$\gamma_{0,1} = 1.5$	1.5014	0.0989	0.0098	1.5014	0.0989	0.0098

从表 5.6.1 中不难看出, 两种方法的均方误差都随着样本量的增加而减小. 注意到, 在 $\sigma = 0.25$ 的时候, 式 (5.6.8) 定义的约束估计量没有明显的改进. 但当 σ 增大到 0.75 的时候, 本节提出的式 (5.6.8) 所定义估计量的均方误差在样本量比较小的时候 ($n = 300$) 改进比较明显. 主要原因是本节所提出的方法在估计 $g_Y(t)$ 和 $g_X(t)$ 时避免了使用估计样本 $\{(\hat{Y}_i, \hat{X}_i, T_i), i = 1, \cdots, n\}$, 而是直接使用观测样本 $\{(\widetilde{Y}_i, \widetilde{X}_i, T_i), i = 1, \cdots, n\}$. 这种改进的估计方法, 从本例中的结果不难看出, 其改进效果在样本量比较小的时候尤为明显.

此外, 还考察了限制约束估计 $\hat{\beta}_R$ 和 $\hat{\gamma}_R$ 的估计效果. 考虑如下两种限制约束条件 (I): $-1.5\beta_{0,1} + \gamma_{0,1} = 0$与限制约束条件 (II): $4\beta_{0,1} + \beta_{0,2} - 2\gamma_{0,1} = 0$ 下的估计

情况. 模拟结果见表 5.6.2.

将表 5.6.2 与表 5.6.1 中的模拟结果做对比, 可看出在限制约束条件 (I)~(II) 下, 式 (5.6.11) 所定义的估计在均方误差意义下得到了明显的改善. 尤其是对参数 $\gamma_{0,1}$ 的均方误差的改进, 在 $\sigma=0.25$ 的时候, 效果尤为明显.

表 5.6.2　模型 (5.6.15) 中参数在两种限制约束条件下估计的均值 (Mean)、标准差 (SD) 和均方误差 (MSE)

σ		n		$\beta_{0,1}=1$	$\beta_{0,2}=-1$	$\gamma_{0,1}=1.5$
$\sigma=0.25$	(I)	300	Mean	1.0110	−0.9888	1.5165
			SD	0.0306	0.0415	0.0460
			MSE	0.0011	0.0018	0.0024
		500	Mean	1.0055	−0.9949	1.5083
			SD	0.0188	0.0280	0.0282
			MSE	0.0004	0.0008	0.0009
		1000	Mean	1.0017	−0.9990	1.5026
			SD	0.0108	0.0168	0.0162
			MSE	0.0001	0.0003	0.0003
	(II)	300	Mean	1.0102	−0.9890	1.5260
			SD	0.0298	0.0410	0.0580
			MSE	0.0010	0.0018	0.0040
		500	Mean	1.0051	−0.9951	1.5126
			SD	0.0184	0.0275	0.0349
			MSE	0.0004	0.0008	0.0014
		1000	Mean	1.0017	−0.9989	1.5040
			SD	0.0105	0.0166	0.0203
			MSE	0.0001	0.0003	0.0004
$\sigma=0.75$	(I)	300	Mean	1.1081	−0.9940	1.5271
			SD	0.0479	0.0657	0.0718
			MSE	0.0026	0.0044	0.0059
		500	Mean	1.0068	−0.9973	1.5102
			SD	0.0322	0.0436	0.0482
			MSE	0.0011	0.0019	0.0024
		1000	Mean	1.0031	−0.9992	1.5047
			SD	0.0226	0.0268	0.0339
			MSE	0.0005	0.0007	0.0012
	(II)	300	Mean	1.0170	−0.9946	1.5366
			SD	0.0465	0.0652	0.0938
			MSE	0.0024	0.0043	0.0101
		500	Mean	1.0062	−0.9978	1.5135
			SD	0.0311	0.0432	0.0643
			MSE	0.0010	0.0019	0.0043
		1000	Mean	1.0030	−0.9993	1.5063
			SD	0.0221	0.0264	0.0443
			MSE	0.0005	0.0007	0.0020

下面考虑如下的假设检验问题:

$$H_0 : \gamma_{0,1} = 0 \quad \text{vs} \quad H_1 : \gamma_{0,1} = c, \tag{5.6.16}$$

其中 c 的取值为 $c = \pm 0.1, \pm 0.2, \cdots, \pm 0.6$. 该检验问题 (5.6.16) 可以看成是协变量 Z 的变量选择问题, 若拒绝原假设 H_0, 变量 Z 对不可观测的响应变量 Y 有影响. 这里参数 $(\beta_{0,1}, \beta_{0,2})^{\mathrm{T}} = (1, -1)^{\mathrm{T}}$, 模型误差 ε 的方差设定为 $\sigma^2 = 0.5^2$. 表 5.6.3 的模拟结果表明, 在原假设 H_0 成立下, wild-bootstrap 方法能够比较准确地刻画犯第一类错误的概率. 随着 c 的绝对值的增加, wild-bootstrap 方法的功效函数能逐渐地趋于 1.

表 5.6.3 假设检验 (5.6.16) 在 c 不同取值下的 wild-bootstrap 功效值

检验水平	$\alpha = 0.01$	$\alpha = 0.025$	$\alpha = 0.05$	$\alpha = 0.10$
$c = 0.00$	0.0101	0.0247	0.0491	0.1017
$c = -0.10$	0.0327	0.0482	0.0905	0.2234
$c = -0.20$	0.2254	0.2599	0.3426	0.4323
$c = -0.30$	0.5113	0.5895	0.6976	0.7807
$c = -0.40$	0.7720	0.8409	0.8648	0.9344
$c = -0.50$	0.9102	0.9357	0.9719	0.9777
$c = -0.60$	0.9689	0.9849	1.0000	1.0000
$c = 0.10$	0.0416	0.1157	0.1745	0.2611
$c = 0.20$	0.1701	0.2578	0.3633	0.4783
$c = 0.30$	0.4270	0.5014	0.5923	0.6901
$c = 0.40$	0.7190	0.7846	0.8345	0.9112
$c = 0.50$	0.8847	0.9379	0.9697	0.9789
$c = 0.60$	0.9607	0.9801	0.9991	1.0000

第 6 章　变系数测量误差模型

本章考虑变系数测量误差模型和部分线性变系数测量误差模型的估计方法、变量选择和统计推断问题, 并研究估计量和检验统计量的渐近性质.

6.1　引　言

6.1.1　变系数模型及研究现状

设 $\{(Y_i, Z_i, T_i), i = 1, \cdots, n\}$ 为含有 n 个个体的样本, 则变系数模型的一般形式为:

$$Y_i = Z_i^{\mathrm{T}} \alpha(T_i) + \varepsilon_i, \quad i = 1, \cdots, n, \tag{6.1.1}$$

其中 $\alpha(\cdot) = (\alpha_1(\cdot), \cdots, \alpha_q(\cdot))^{\mathrm{T}}$ 为 $q \times 1$ 未知函数系数向量, Z_i 和 T_i 为协变量, Y_i 为响应变量, ε_i 为模型误差且满足 $E(\varepsilon_i|Z_i, T_i) = 0$. 若 $Z_{i1} \equiv 1$, 则意味着模型 (6.1.1) 中包含变截距项. 变系数模型之所以有吸引力是因为它允许模型中的回归系数为非参数函数, 即系数 $\alpha_j(T)$ 依赖于协变量 T, 这不但削减了建模偏差提高了模型适应性, 而且避免了 "维数祸根" 问题; 同时, 由于它形式上仍具有线性结构, 所以在实际应用中容易理解和解释. 因此该模型在生物医学、计量经济学、环境科学等诸多领域都有着广泛的应用, 是处理复杂数据的一个有力工具.

自从 Cleveland 等 (1992) 与 Hastie 和 Tibshirani (1993) 率先对变系数模型进行研究以来, 许多统计学者对该模型进行了深入研究, 并取得了丰富的研究成果. 在独立数据下, Fan 和 Zhang (1999) 提出了系数函数 $\alpha(\cdot)$ 的两阶段估计过程, 并基于局部多项式方法构造了 $\alpha(\cdot)$ 的估计量, 给出了所得估计量的渐近均方误差及最优收敛速度. Fan 和 Zhang (2000) 研究了模型中系数函数 $\alpha(\cdot)$ 的同时置信带构造及假设检验问题. Cai 等 (2000) 基于局部多项式方法研究了广义变系数模型的估计和检验问题, 证明了所得系数函数估计的渐近正态性, 并采用非参数极大似然比检验方法检验模型中的系数函数是否为常数. Zhang 和 Lee (2000) 研究了系数函数 $\alpha(\cdot)$ 的局部多项式估计的变窗宽选择问题. Cai (2002) 利用两阶段局部似然方法讨论了 $\alpha(\cdot)$ 的估计, 并证明了所得估计达到了非参数最优收敛速度. Eubank 等 (2004) 利用光滑样条方法对 $\alpha(\cdot)$ 的估计问题进行了研究. Lu 和 Mao (2004) 基于 B 样条方法构造了系数函数 $\alpha(\cdot)$ 的估计. 张日权和卢一强 (2004) 对变系数模型进行了进一

步研究. Ip 等 (2007) 将广义似然比检验方法应用于系数函数的检验, 证明了所提出的检验统计量渐近服从 χ^2 分布. Leng (2009) 与 Hu 和 Xia (2012) 分别利用惩罚估计的方法来识别变系数模型中的常系数和变系数效应分量. Wang 和 Xia (2009) 结合局部多项式光滑和惩罚估计思想研究了模型的变量选择问题. Tang 等 (2012) 对模型 (6.1.1) 提出一个可以同时识别常系数和变系数效应分量并选择显著变量的方法. Lian (2012) 基于多项式样条展开和 adaptive group Lasso 估计方法研究了广义变系数模型的变量选择问题.

此外, 在纵向数据情形下, Wu 等 (1998) 基于核光滑和最小二乘准则得到了系数函数 $\alpha(\cdot)$ 的估计并证明其渐近性质, 进一步采用所得结果构造了系数函数 $\alpha(\cdot)$ 的置信域和同时置信带. Hoover 等 (1998) 分别利用光滑样条方法及加权局部多项式方法研究了 $\alpha(\cdot)$ 的估计问题. Wu 等 (2000) 提出先中心化协变量, 然后采用局部最小二乘方法给出了函数系数的两步核估计, 并证明了估计量的渐近性质. Huang 等 (2002) 利用基函数逼近思想, 研究了系数函数 $\alpha(\cdot)$ 的整体光滑估计方法. Huang 等 (2004) 采用多项式样条方法逼近每一个系数函数, 然后基于最小二乘准则给出了函数系数的估计, 并证明了所得估计量的相合性、收敛速度及渐近正态性. Qu 和 Li (2006) 采用二次推断函数方法研究了变系数模型的估计和检验问题. Xue 和 Zhu (2007) 利用经验似然方法, 提出了系数函数 $\alpha(\cdot)$ 的两种纠偏的经验似然比, 证明了所构造的经验对数似然比渐近于标准 χ^2 分布, 并且利用所得结果构造了系数函数分量的逐点置信区间和同时置信带. Wang 等 (2008) 基于基函数展开和惩罚估计方法考虑了带有重复测量数据情形的变系数模型的变量选择问题. Wei 等 (2011) 基于 group Lasso 和基函数逼近方法研究了高维变系数模型的变量选择问题. 更多关于变系数模型 (6.1.1) 的研究成果可参见文献 Wu 和 Chiang (2000), Chiang 等 (2001), Cai (2003), Honda (2004), Kauermann 等 (2005), Tang 和 Wang (2005a, 2005b), Kim (2007), Tang 和 Cheng (2008), Lu 等 (2008), Wang (2009), Wang 和 Zhang (2009), Zhang 和 Peng (2010), Li 等 (2011c), Wei (2011), Homg 和 Lian (2012), Jiang 等 (2013), Klopp 和 Pensky (2013), Feng 等 (2015) 等.

6.1.2 部分线性变系数模型及研究现状

Zhang 等 (2002) 在实际应用中发现模型 (6.1.1) 中的部分系数分量是常数系数而不是函数系数, 因此将变系数模型推广为部分线性变系数模型, 其形式如下:

$$Y_i = X_i^{\mathrm{T}}\beta + Z_i^{\mathrm{T}}\alpha(T_i) + \varepsilon_i, \quad i = 1, \cdots, n, \tag{6.1.2}$$

其中 Y_i 是响应变量, X_i, Z_i 和 T_i 为协变量, $\alpha(\cdot) = (\alpha_1(\cdot), \cdots, \alpha_q(\cdot))^{\mathrm{T}}$ 为 q 维未知系数函数向量, $\beta = (\beta_1, \cdots, \beta_p)^{\mathrm{T}}$ 为 p 维未知参数向量, ε_i 是相互独立的随机误差且满足 $E(\varepsilon_i|X_i, Z_i, T_i) = 0$. 显然, 模型 (6.1.2) 是经典线性模型和变系数模型的推

广, 它包含了许多重要的统计模型. 例如, 当 $\alpha(\cdot) = \alpha$(常数向量) 时, 模型 (6.1.2) 即为传统的线性模型. 当 $q = 1$ 且 $Z = 1$ 时, 模型 (6.1.2) 退化为部分线性模型. 当 $X = 0$ 时, 该模型即为经典的变系数模型.

部分线性变系数模型由于其融合了传统线性模型和变系数模型的优点, 具有更好的模型适应性和灵活性, 因此近年来得到了广泛的关注. Zhang 等 (2002) 基于局部多项式方法研究了部分线性变系数模型中参数分量和非参数分量的估计. Xia 等 (2004) 提出了一个新的估计方法, 构造了模型中参数分量的一个有效估计. Zhou 和 You (2004) 利用小波方法估计了模型中的未知参数分量和非参数分量. Fan 和 Huang (2005) 基于局部线性展开和 profile 最小二乘方法得到了模型中参数分量的估计, 证明了所得估计量的渐近性质, 并进一步利用 profile 广义似然比检验方法考虑了模型的检验问题. Ahmad 等 (2005) 采用广义级数方法构造了模型 (6.1.2) 中未知参数分量的估计, 并证明了所得估计的相合性和渐近正态性. You 和 Zhou (2006) 研究了模型中参数分量的经验似然推断问题. Huang 和 Zhang (2009) 利用经验似然方法研究了模型中非参数分量的统计推断问题. Li 等 (2012a) 研究了参数分量维数发散的部分线性变系数模型经验似然推断问题. Li 和 Liang (2008) 结合惩罚估计以及半参数广义似然比检验方法研究了广义部分线性变系数模型中参数分量和非参数分量的变量选择问题. Zhao 和 Xue (2009) 基于基函数展开和惩罚最小二乘方法, 对模型 (6.1.2) 提出一个可以同时对参数分量和非参数分量进行变量选择的方法. Wang 等 (2009) 利用假设检验方法研究了部分线性变系数分位数回归的模型选择问题, 并对基函数逼近和惩罚方法的变量选择过程进行了讨论. Kai 等 (2011) 利用复合分位数回归, 提出一种新的有效估计和变量选择方法. Gu 和 Liang (2014) 研究了部分线性变系数模型的协整关系检验问题. 关于部分线性变系数模型的其他研究工作可以参见文献 Lam 和 Fan (2008), Lu (2008), Hu 等 (2009), Zhou 和 Liang (2009), Li 等 (2009), Hu 和 Cui (2010), Yang 和 Li (2010), Huang (2011), Li 等 (2011a), Zhang 等 (2013c), Shao 等 (2014), Bai 等 (2015) 等.

6.2 变系数测量误差模型的局部偏差校正统计推断

本节考虑协变量 Z 具有可加测量误差, 且协变量 T 准确测量的变系数 EV 模型. 假设 $\{(W_i = (W_{i1}, \cdots, W_{iq})^{\mathrm{T}}, T_i, Y_i), 1 \leqslant i \leqslant n\}$ 为来自如下模型 (6.2.1) 的一组独立同分布的可观测随机样本, 即考虑如下的变系数 EV 模型:

$$\begin{cases} Y_i = Z_i^{\mathrm{T}} \alpha(T_i) + \varepsilon_i, \\ W_i = Z_i + u_i, \quad i = 1, \cdots, n, \end{cases} \tag{6.2.1}$$

其中协变量 Z_i 带有可加测量误差, $W_i = (W_{i1}, \cdots, W_{iq})^{\mathrm{T}}$ 为 Z_i 的替代变量, ε_i 是均值为 0 方差为 σ^2 的模型误差, u_i 为不可观测的 q 维随机误差向量, u_i 与 $(Z_i^{\mathrm{T}}, T_i, \varepsilon_i)^{\mathrm{T}}$ 相互独立, 均值为 0, 协方差阵为 Σ_{uu}.

6.2.1 局部偏差校正估计

You 等 (2006) 在假定测量误差协方差阵 Σ_{uu} 已知情况下, 讨论了模型 (6.2.1) 中未知系数函数的估计问题. 对于 Σ_{uu} 未知的情形, 可以采用重复测量的方法来对其进行估计, 具体细节见 2.2.1 节.

注意到, 若 Z_i 可以精确观测, 局部线性回归方法可以用来估计系数函数 $\{\alpha_j(\cdot), j = 1, \cdots, q\}$. 在任意给定 t 的某个小邻域内, $\alpha_j(T)$ 可以被如下线性函数进行局部线性逼近:

$$\alpha_j(T) \approx \alpha_j(t) + \alpha_j'(t)(T - t) =: a_j + b_j(T - t), \quad j = 1, \cdots, q, \tag{6.2.2}$$

其中 $\alpha_j'(t) = \partial\alpha_j(t)/\partial t$. 于是, 可以通过最小化如下局部加权最小二乘问题来估计系数函数 $\alpha(T)$:

$$\sum_{i=1}^{n}\{Y_i - Z_i^{\mathrm{T}}[a + b(T_i - t)]\}^2 K_h(T_i - t), \tag{6.2.3}$$

其中 $a = (a_1, \cdots, a_q)^{\mathrm{T}}$, $b = (b_1, \cdots, b_q)^{\mathrm{T}}$, $K_h(\cdot) = K(\cdot/h)/h$, $K(\cdot)$ 为核函数, h 为窗宽.

为描述方便, 记 $Y = (Y_1, \cdots, Y_n)^{\mathrm{T}}$, $W = (W_1, \cdots, W_n)^{\mathrm{T}}$, $Z = (Z_1, \cdots, Z_n)^{\mathrm{T}}$, $\omega_t = \operatorname{diag}(K_h(T_1 - t), \cdots, K_h(T_n - t))$, 以及

$$D_t^Z = \begin{pmatrix} Z_1^{\mathrm{T}} & \dfrac{T_1 - t}{h} Z_1^{\mathrm{T}} \\ \vdots & \vdots \\ Z_n^{\mathrm{T}} & \dfrac{T_n - t}{h} Z_n^{\mathrm{T}} \end{pmatrix}.$$

则基于式 (6.2.3), 由广义最小二乘法可得

$$[\hat{a}^{\mathrm{T}}, h\hat{b}^{\mathrm{T}}]^{\mathrm{T}} = \left\{(D_t^Z)^{\mathrm{T}}\omega_t D_t^Z\right\}^{-1}(D_t^Z)^{\mathrm{T}}\omega_t Y. \tag{6.2.4}$$

现在, 由于 Z_i 不可观测, 观测到的是带有误差扰动的变量 W_i, You 等 (2006) 指出, 若忽略测量误差直接用 W_i 代替式 (6.2.3) 中的 Z_i, 则会导致所得系数函数 $\alpha(\cdot)$ 的估计是不相合的. 为了克服由测量误差所造成的估计的不相合性, You 等

(2006) 提出了局部偏差校正的思想, 主要想法如下. 注意到,

$$\begin{aligned}&\frac{1}{n}(D_t^W)^{\mathrm{T}}\omega_t D_t^W\\&=\frac{1}{n}\sum_{i=1}^n Z_iZ_i^{\mathrm{T}}\otimes\begin{pmatrix}1 & \dfrac{T_i-t}{h}\\ \dfrac{T_i-t}{h} & \left(\dfrac{T_i-t}{h}\right)^2\end{pmatrix}K_h(T_i-t)\\&\quad+\frac{1}{n}\sum_{i=1}^n \Sigma_{uu}\otimes\begin{pmatrix}1 & \dfrac{T_i-t}{h}\\ \dfrac{T_i-t}{h} & \left(\dfrac{T_i-t}{h}\right)^2\end{pmatrix}K_h(T_i-t)+O\left\{\left(\frac{\log n}{nh}\right)^{1/2}\right\}\quad \text{a.s.},\end{aligned}$$

其中 D_t^W 是由 W_i 替换 D_t^Z 中的 Z_i 所得, $\otimes$ 表示 Kronecker 乘积. 记

$$\Omega=\sum_{i=1}^n \Sigma_{uu}\otimes\begin{pmatrix}1 & \dfrac{T_i-t}{h}\\ \dfrac{T_i-t}{h} & \left(\dfrac{T_i-t}{h}\right)^2\end{pmatrix}K_h(T_i-t).$$

You 等 (2006) 定义 $\Phi(t)=(\alpha_1(t),\cdots,\alpha_q(t),\alpha_1'(t),\cdots,\alpha_q'(t))^{\mathrm{T}}$ 的局部偏差校正估计为

$$\hat{\Phi}(t)=H^{-1}\left\{(D_t^W)^{\mathrm{T}}\omega_t D_t^W-\Omega\right\}^{-1}(D_t^W)^{\mathrm{T}}\omega_t Y,$$

其中 $H=\mathrm{diag}\{1,h\}\otimes I_q$, I_q 为 $q\times q$ 单位矩阵. 记 $e_{j,2q}$ 为第 j 个元素为 1, 其他元素均为 0 的 $2q$ 维列向量, 则 $\alpha_j(t)$ 的估计为

$$\hat{\alpha}_j^W(t)=e_{j,2q}^{\mathrm{T}}H^{-1}\left\{(D_t^W)^{\mathrm{T}}\omega_t D_t^W-\Omega\right\}^{-1}(D_t^W)^{\mathrm{T}}\omega_t Y,$$

并称 $\hat{\alpha}^W(t)=(\hat{\alpha}_1^W(t),\cdots,\hat{\alpha}_q^W(t))^{\mathrm{T}}$ 为系数函数 $\alpha(t)$ 的局部校正估计.

进一步, 由 $E(Y_i-Z_i^{\mathrm{T}}\alpha(T_i))^2=\sigma^2$ 和 $E(Y_i-W_i^{\mathrm{T}}\alpha(T_i))^2=\sigma^2+E(\alpha^{\mathrm{T}}(T_1)\Sigma_{uu}\alpha(T_1))$, 可以定义模型误差方差 σ^2 的估计如下:

$$\hat{\sigma}_n^2=\frac{1}{n}\sum_{i=1}^n(Y_i-W_i^{\mathrm{T}}\hat{\alpha}^W(T_i))^2-\frac{1}{n}\sum_{i=1}^n(\hat{\alpha}^W(T_i))^{\mathrm{T}}\Sigma_{uu}\hat{\alpha}^W(T_i).$$

为了得到估计量的渐近结果, 需要下面的正则条件.

(C6.2.1) $(Z_i^{\mathrm{T}},T_i,\varepsilon_i,u_i^{\mathrm{T}})^{\mathrm{T}}$ 为独立同分布随机向量. $(Z_i^{\mathrm{T}},T_i)^{\mathrm{T}}$ 与 $(\varepsilon_i,u_i^{\mathrm{T}})^{\mathrm{T}}$ 相互独立, 且 ε_i 和 u_i 独立. T_1 有有界支撑 $\mathcal{T}$, 其密度函数 $f_T(t)$ 是 Lipschitz 连续的且在 $\mathcal{T}$ 上有界远离 0.

(C6.2.2) 对每一个 $t\in\mathcal{T}$, 矩阵 $\Gamma(t)=E(Z_1Z_1^{\mathrm{T}}|T_1=t)$ 非退化.

(C6.2.3) 存在常数 $s>2$ 使得 $E\varepsilon_1^{2s}<\infty$, $E\|u_1\|^{2s}<\infty$, $E\|Z_1\|^{2s}<\infty$. 对某个 $\delta<2-s^{-1}$, 当 $n\to\infty$ 时, 有 $n^{2\delta-1}h\to\infty$.

(C6.2.4) $\{\alpha_j(\cdot), j=1,\cdots,q\}$ 在 $\mathcal{T}$ 上有连续的二阶导数.

(C6.2.5) 核函数 $K(\cdot)$ 为概率密度函数且有紧支撑. 当 $n\to\infty$ 时, $h\to 0$, $nh^8\to 0$, $nh^2/(\log n)^2\to\infty$.

令$\mu_j=\int t^jK(t)\mathrm{d}t$, $\nu_j=\int t^jK^2(t)\mathrm{d}t$, $j=0,1,2,3$. You 等 (2006) 给出了下面几个定理及推论研究系数函数的估计 $\hat{\alpha}^W(t)$ 和误差方差估计 $\hat{\sigma}_n^2$ 的渐近性质. 定理的详细证明, 请读者参考 You 等 (2006).

定理 6.2.1 假设条件(C6.2.1)~条件(C6.2.5)成立, 当 $n\to\infty$ 时, 则有

$$\sqrt{nh}\left[H\left\{\hat{\Phi}(t)-\Phi(t)\right\}-\frac{h^2}{2}\frac{1}{\mu_2-\mu_1^2}\begin{pmatrix}(\mu_2-\mu_1\mu_3)\alpha''(t)\\(\mu_3-\mu_1\mu_2)\alpha''(t)\end{pmatrix}+o(h^2)\right]\xrightarrow{L}N(0,\Sigma),$$

其中 $\alpha''(t)=(\alpha_1''(t),\cdots,\alpha_q''(t))^{\mathrm{T}}$, $\alpha_j''(t)=\partial^2\alpha_j(t)/\partial t^2$,

$$\begin{aligned}\Sigma=&\frac{1}{f_T(t)}\Sigma^*\otimes\frac{1}{(\mu_2-\mu_1^2)^2}\\&\times\begin{pmatrix}\mu_2^2\nu_0-2\mu_1\mu_2\nu_1+\mu_1^2\nu_2 & (\mu_1^2+\mu_2)\nu_1-\mu_1\mu_2\nu_0-\mu_1\nu_2\\(\mu_1^2+\mu_2)\nu_1-\mu_1\mu_2\nu_0-\mu_1\nu_2 & \nu_2-\mu_1(2\nu_1+\mu_1\nu_0)\end{pmatrix},\end{aligned}$$

$$\Sigma^*=\Gamma^{-1}(t)[\sigma^2\Gamma(t)+\sigma^2\Sigma_{uu}+E\{\xi_1\alpha(t)\alpha^{\mathrm{T}}(t)\xi_1^{\mathrm{T}}|T_1=t\}]\Gamma^{-1}(t),$$

$$\xi_1=\Sigma_{uu}-u_1u_1^{\mathrm{T}}-Z_1u_1^{\mathrm{T}}.$$

推论 6.2.1 假设定理 6.2.1 的条件成立, 当 $n\to\infty$ 时, 则有

$$\sqrt{nh}\left\{\hat{\alpha}^W(t)-\alpha(t)-\frac{h^2}{2}\frac{\mu_2^2-\mu_1\mu_3}{\mu_2-\mu_1^2}\alpha''(t)+o(h^2)\right\}\xrightarrow{L}N\left(0,\Sigma^W\right),$$

其中 $\Sigma^W=(c_0^2\nu_0+2c_0c_1\nu_1+c_1^2\nu_2)\Sigma^*/f_T(t)$, $c_0=\mu_2/(\mu_2-\mu_1^2)$, $c_1=-\mu_1/(\mu_2-\mu_1^2)$.

为应用推论 6.2.1 进行统计推断, 需要估计 $\hat{\alpha}^W(\cdot)$ 的渐近方差. 由于 μ_j 和 ν_j 为已知常数, 因此只需给出 $\Sigma^*/f_T(t)$ 的相合估计即可. 定义

$$\hat{\Gamma}(t)=\frac{1}{n}\sum_{i=1}^nK_h(T_i-t)Z_iZ_i^{\mathrm{T}},$$

$$G_i=W_i(Y_i-W_i^{\mathrm{T}}\hat{\alpha}^W(T_i))+\Sigma_{uu}\hat{\alpha}^W(T_i),$$

$$\hat{\Phi}(t)=\frac{1}{n}\sum_{i=1}^nK_h(T_i-t)G_iG_i^{\mathrm{T}}.$$

下面的定理说明 $\{\hat{\Gamma}(t)\}^{-1}\hat{\Phi}(t)\{\hat{\Gamma}(t)\}^{-1}$ 是 $\Sigma^*/f_T(t)$ 的相合估计.

定理 6.2.2　假设条件(C6.2.1)~条件(C6.2.5)成立, 当 $n \to \infty$ 时, 则有

$$\{\hat{\Gamma}(t)\}^{-1}\hat{\Phi}(t)\{\hat{\Gamma}(t)\}^{-1} \xrightarrow{P} \Sigma^*/f_T(t).$$

下面的定理给出了 $\hat{\alpha}_j^W(t)$ 的一致强收敛速度.

定理 6.2.3　假设条件(C6.2.1)~条件(C6.2.5)成立, 当 $n \to \infty$ 时, 则有

$$\max_{1\leqslant j\leqslant q}\sup_{t\in\mathcal{T}}|\hat{\alpha}_j^W(t)-\alpha_j(t)| = O\{h^2+(\log n/nh)^{1/2}\} \quad \text{a.s..}$$

注 6.2.1　若 h 取最优窗宽, 即 $h = cn^{-1/5}$, 其中 c 为一个常数, 则由定理 6.2.3 可知

$$\max_{1\leqslant j\leqslant q}\sup_{t\in\mathcal{T}}|\hat{\alpha}_j^W(t)-\alpha_j(t)| = O\left\{n^{-2/5}(\log n)^{1/2}\right\} \quad \text{a.s..}$$

这说明系数函数的局部偏差校正估计的收敛速度达到了非参数函数估计的最优一致强收敛速度 (Stone, 1982).

定理 6.2.4　假设条件(C6.2.1)~条件(C6.2.5)成立, 当 $n \to \infty$ 时, 则有

$$\sqrt{n}(\hat{\sigma}_n^2-\sigma^2) \xrightarrow{L} N(0,\Delta),$$

其中 $\Delta = \mathrm{Var}\left\{\varepsilon_1^2 - \alpha^{\mathrm{T}}(T_1)\left(u_1u_1^{\mathrm{T}}-\Sigma_{uu}\right)\alpha(T_1) - 2\varepsilon_1u_1\alpha(T_1)\right\}$.

定理 6.2.5　假设条件(C6.2.1)~条件(C6.2.5), 当 $n \to \infty$ 时, 则有

$$\lim_{n\to\infty}\sqrt{\frac{n}{2\log\log n}}|\hat{\sigma}_n^2-\sigma^2| \leqslant \sqrt{\Delta} \quad \text{a.s..}$$

6.2.2　局部纠偏经验似然

根据 6.2.1 节提出的估计方法及估计量的渐近正态性可以构造系数函数的逐点置信区间, 但需要相合地估计渐近偏差和渐近方差, 这可能会影响置信区间构造的精度. 为了避免该问题, 冯三营等 (2010b) 采用经验似然方法研究了模型 (6.2.1), 提出了局部纠偏的经验似然比函数, 由此构造系数函数的逐点置信区间. 下面对冯三营等 (2010b) 提出的经验似然方法进行简单介绍.

若无测量误差, 即 Z_i 已知, 记 $\theta = (a^{\mathrm{T}}, b^{\mathrm{T}})^{\mathrm{T}}$, 则引入辅助随机向量

$$\widetilde{\eta}_i(\theta) = \left(Z_i^{\mathrm{T}}, Z_i^{\mathrm{T}}(T_i-t)\right)^{\mathrm{T}}\left[Y_i - \left(Z_i^{\mathrm{T}}, Z_i^{\mathrm{T}}(T_i-t)\right)\theta\right]K_h(T_i-t). \tag{6.2.5}$$

显然, $E(\widetilde{\eta}_i(\theta)) = o(1)$, 从而定义经验对数似然比函数

$$\widetilde{\mathcal{R}}(\theta) = -2\max\left\{\sum_{i=1}^n \log(np_i)\,\middle|\, p_i \geqslant 0,\ \sum_{i=1}^n p_i = 1,\ \sum_{i=1}^n p_i\widetilde{\eta}_i(\theta) = 0\right\}. \tag{6.2.6}$$

在适当的条件下可以证明 $\widetilde{\mathcal{R}}(\theta) \xrightarrow{L} \chi^2_{2q}$. 由此可以构造 θ 的置信域. 再由式 (6.2.2) 即可得到系数函数 $\alpha(T_i)$ 的置信域.

现在由于 Z_i 不可观测, 观测到的是带有误差扰动的变量 W_i, 因此引入如下局部纠偏的辅助随机向量

$$\begin{aligned}\eta_i(\theta) = &\left(W_i^{\mathrm{T}}, W_i^{\mathrm{T}}(T_i - t)\right)^{\mathrm{T}} \left[Y_i - \left(W_i^{\mathrm{T}}, W_i^{\mathrm{T}}(T_i - t)\right)\theta\right] K_h(T_i - t)\\ &+ C_i\theta K_h(T_i - t),\end{aligned} \tag{6.2.7}$$

其中 $C_i = \{(1, T_i - t)^{\mathrm{T}}(1, T_i - t)\} \otimes \Sigma_{uu}$. 式 (6.2.7) 最后一项的作用是修正由测量误差引起的偏差. 因此, 定义经验对数似然比函数

$$\mathcal{R}(\theta) = -2\max\left\{\sum_{i=1}^{n}\log(np_i)\middle|\, p_i \geqslant 0,\ \sum_{i=1}^{n}p_i = 1,\ \sum_{i=1}^{n}p_i\eta_i(\theta) = 0\right\}. \tag{6.2.8}$$

由 Lagrange 乘子法, $\mathcal{R}(\theta)$ 可以表示为

$$\mathcal{R}(\theta) = 2\sum_{i=1}^{n}\log\{1 + \lambda^{\mathrm{T}}\eta_i(\theta)\}, \tag{6.2.9}$$

其中 λ 是 Lagrange 乘子, 满足

$$\frac{1}{n}\sum_{i=1}^{n}\frac{\eta_i(\theta)}{1 + \lambda^{\mathrm{T}}\eta_i(\theta)} = 0.$$

极小化 $\mathcal{R}(\theta)$ 可得 θ 的估计 $\hat{\theta}$, 称为极大经验似然估计 (MELE). 由式 (6.2.9) 及 Taylor 展开易证

$$\mathcal{R}(\theta) = \left(\frac{1}{\sqrt{nh}}\sum_{i=1}^{n}\eta_i(\theta)\right)^{\mathrm{T}} S^{-1}\left(\frac{1}{\sqrt{nh}}\sum_{i=1}^{n}\eta_i(\theta)\right) + o_P(1), \tag{6.2.10}$$

其中 $S = \dfrac{1}{nh}\sum\limits_{i=1}^{n}\eta_i(\theta)\eta_i^{\mathrm{T}}(\theta)$, 类似于 Xue 和 Zhu (2007) 中式 (2.6) 的讨论可得

$$\hat{\theta} = \left\{\sum_{i=1}^{n}K_h(T_i - t)(\zeta_i\zeta_i^{\mathrm{T}} - C_i)\right\}^{-1}\left(\sum_{i=1}^{n}K_h(T_i - t)\zeta_i Y_i\right) + o_P\left((nh)^{-1/2}\right), \tag{6.2.11}$$

其中 $\zeta_i = \left(W_i^{\mathrm{T}}, W_i^{\mathrm{T}}(T_i - t)\right)^{\mathrm{T}}$.

通常情况下, 人们对参数 θ 的第一部分, 即 a 的置信域感兴趣, 为此定义

$$\hat{\eta}_i(a) = \left(I_q\ 0_q\right)\left\{\frac{1}{nh}\sum_{i=1}^{n}K_h(T_i - t)(\zeta_i\zeta_i^{\mathrm{T}} - C_i)\right\}^{-1}\eta_i((a^{\mathrm{T}}, \hat{b}^{\mathrm{T}})^{\mathrm{T}}), \tag{6.2.12}$$

其中 $\hat{b} = (0_q\ I_q)\hat{\theta}$, I_q 为 q 阶单位阵, 0_q 为 q 阶全零矩阵. 类似于式 (6.2.8), 可以定义经验对数似然比函数

$$\hat{\mathcal{R}}(a) = -2\max\left\{\sum_{i=1}^{n}\log(np_i)\ \middle|\ p_i \geqslant 0,\ \sum_{i=1}^{n}p_i = 1,\ \sum_{i=1}^{n}p_i\hat{\eta}_i(a) = 0\right\}, \tag{6.2.13}$$

由 Lagrange 乘子法, $\hat{\mathcal{R}}(a)$ 可以表示为

$$\hat{\mathcal{R}}(a) = 2\sum_{i=1}^{n}\log\{1 + \lambda_1^{\mathrm{T}}\hat{\eta}_i(a)\}, \tag{6.2.14}$$

其中 λ_1 是 Lagrange 乘子, 满足

$$\frac{1}{n}\sum_{i=1}^{n}\frac{\hat{\eta}_i(a)}{1 + \lambda_1^{\mathrm{T}}\hat{\eta}_i(a)} = 0.$$

定理 6.2.6　假设条件(C6.2.1)～条件(C6.2.5)成立, 若 θ 是参数真值, 当 $n \to \infty$ 时, 则有 $\mathcal{R}(\theta) \xrightarrow{L} \chi_{2q}^2$.

基于定理 6.2.6, 可以定义参数向量 θ 的置信域, 即对任给的 $0 < \delta < 1$, 存在 c_δ 使得 $P\{\chi_{2q}^2 \leqslant c_\delta\} = 1 - \delta$, 则

$$I_\delta(\theta^*) = \left\{\theta^*\ \middle|\ \mathcal{R}(\theta^*) \leqslant c_\delta\right\}$$

为参数向量 θ 的具有渐近置信水平 $1-\delta$ 的置信域, 并且有 $P(\theta \in I_\delta(\theta^*)) = 1 - \delta + o(1)$. 结合式 (6.2.2) 可以得到 $\alpha(\cdot)$ 的渐近置信水平 $1-\delta$ 的逐点置信域.

定理 6.2.7　假设条件(C6.2.1)～条件(C6.2.5)成立, 若 a 是参数真值, 当 $n \to \infty$ 时, 则有 $\hat{\mathcal{R}}(a) \xrightarrow{L} \chi_q^2$.

基于定理 6.2.7, 可以使用 $\hat{\mathcal{R}}(a)$ 构造 a 的置信域

$$\hat{I}_\delta(a^*) = \left\{a^*\ \middle|\ \hat{\mathcal{R}}(a^*) \leqslant d_\delta\right\},$$

这里 $P(\chi_q^2 \leqslant d_\delta) = 1 - \delta$, 并且有 $P(a \in \hat{I}_\delta(a^*)) = 1 - \delta + o(1)$. 从而, 结合式 (6.2.2) 也可以得到 $\alpha(\cdot)$ 的渐近置信水平 $1-\delta$ 的逐点置信域.

6.3 变系数测量误差模型的变量选择

本节考虑模型 (6.2.1) 的变量选择问题. Zhao 和 Xue (2011) 利用基函数逼近、偏差校正和惩罚最小二乘方法, 对模型 (6.2.1) 提出了一种变量选择方法. 通过选择合适的截断参数, 证明了所提出的变量选择方法可以相合地识别出真实的模型, 且所得回归系数的正则估计具有 oracle 性质. 具体方法和结论是本节的主要内容.

6.3.1 估计方法

设 $B(u)=(B_1(u),\cdots,B_L(u))^{\mathrm{T}}$ 为 $M+1$ 阶 B 样条基函数, 其中 $L=K+M+1$, K 为内节点个数, 那么, $\alpha_k(t)$ 可以由下式逼近:

$$\alpha_k(t)\approx B^{\mathrm{T}}(t)\beta_k,\quad k=1,\cdots,q. \tag{6.3.1}$$

若 Z_i 可观测, 则将式 (6.3.1) 代入模型 (6.1.1) 可得

$$Y_i\approx V_i^{\mathrm{T}}\beta+\varepsilon_i,\quad i=1,\cdots,n,$$

其中 $V_i=I_q\otimes B(T_i)\cdot Z_i$, $\beta=(\beta_1^{\mathrm{T}},\cdots,\beta_q^{\mathrm{T}})^{\mathrm{T}}$. 从而, 可以定义如下惩罚最小二乘目标函数:

$$\sum_{i=1}^{n}(Y_i-V_i^{\mathrm{T}}\beta)^2+n\sum_{k=1}^{q}p_\lambda(\|\beta_k\|_H), \tag{6.3.2}$$

其中 $\|\beta_k\|_H=(\beta_k^{\mathrm{T}}H\beta_k)^{1/2}$, $H=\displaystyle\int B(t)B^{\mathrm{T}}(t)\mathrm{d}t$, $p_\lambda(\cdot)$ 为惩罚函数, λ 为截断参数. 此处采用 SCAD 惩罚函数:

$$p'_\lambda(\omega)=\lambda\left\{I(\omega\leqslant\lambda)+\frac{(a\lambda-\omega)_+}{(a-1)\lambda}I(\omega>\lambda)\right\},$$

其中 $a>2$, $\omega>0$ 且 $p_\lambda(0)=0$.

现在, 由于模型 (6.2.1) 中的 Z_i 不可观测, 若忽略测量误差直接用 W_i 代替 Z_i 则可以证明所得正则估计是不相合的. 另外, You 等 (2006) 提出的基于局部多项式估计的纠偏方法对于这里的基函数展开方法也是不适用的. 为此, Zhao 和 Xue (2011) 提出一个新的纠偏估计方法. 具体地, 记 $\widetilde{V}_i=I_q\otimes B(T_i)\cdot W_i$, 则简单计算可得

$$E\left(\widetilde{V}_i\widetilde{V}_i^{\mathrm{T}}\big|Z_i,T_i\right)=V_iV_i^{\mathrm{T}}+\Omega(T_i),$$

其中 $\Omega(t)=[I_q\otimes B(t)]\Sigma_{uu}[I_q\otimes B(t)]^{\mathrm{T}}$. 因此, 定义纠偏惩罚最小二乘目标函数为

$$Q(\beta)=\sum_{i=1}^{n}(Y_i-\widetilde{V}_i^{\mathrm{T}}\beta)^2-\sum_{i=1}^{n}\beta^{\mathrm{T}}\Omega(T_i)\beta+n\sum_{k=1}^{q}p_\lambda(\|\beta_k\|_H), \tag{6.3.3}$$

其中等号右端第二项是为了消除 $\widetilde{V}_i$ 中的测量误差带来的影响. 记 $\hat{\beta}=(\hat{\beta}_1^{\mathrm{T}},\cdots,\hat{\beta}_q^{\mathrm{T}})^{\mathrm{T}}$ 为最小化式 (6.3.3) 的解, 则 $\alpha_k(t)$ 的估计为 $\hat{\alpha}_k(t)=B^{\mathrm{T}}(t)\hat{\beta}_k$.

6.3.2 算法

因为 $Q(\beta)$ 中的惩罚函数在原点是奇异的, 所以普通的梯度算法将不能直接应用. 类似于 Fan 和 Li (2001), 基于惩罚函数 $p_\lambda(\cdot)$ 的局部二次逼近, 给出一个迭代计算方法. 具体地, 在任一给定非零 ω_0 的某个邻域内, 惩罚函数 $p_\lambda(\cdot)$ 在 ω_0 点可以近似表示为

$$p_\lambda(|\omega|)\approx p_\lambda(|\omega_0|)+\frac{1}{2}\frac{p_\lambda'(|\omega_0|)}{|\omega_0|}(\omega^2-\omega_0^2).$$

于是, 对给定的初始值 $\beta_k^{(0)}$ $(\|\beta_k^{(0)}\|_H>0)$, $k=1,\cdots,q$, 有

$$p_\lambda(\|\beta_k\|_H)\approx p_\lambda(\|\beta_k^{(0)}\|_H)+\frac{1}{2}\frac{p_\lambda'(\|\beta_k^{(0)}\|_H)}{\|\beta_k^{(0)}\|_H}(\|\beta_k\|_H^2-\|\beta_k^{(0)}\|_H^2).$$

记

$$\Sigma_\lambda(\beta^{(0)})=\mathrm{diag}\left\{\frac{p_\lambda'(\|\beta_1^{(0)}\|_H)}{\|\beta_1^{(0)}\|_H}H,\cdots,\frac{p_\lambda'(\|\beta_q^{(0)}\|_H)}{\|\beta_q^{(0)}\|_H}H\right\},$$

于是, 除了一个常数项, 式 (6.3.3) 可写为

$$Q(\beta)=\sum_{i=1}^n(Y_i-\widetilde{V}_i^{\mathrm{T}}\beta)^2-\sum_{i=1}^n\beta^{\mathrm{T}}\Omega(T_i)\beta+\frac{n}{2}\beta^{\mathrm{T}}\Sigma_\lambda(\beta^{(0)})\beta. \tag{6.3.4}$$

注意到, 上式是一个二次型, 并且可以通过下式求解:

$$\left(\sum_{i=1}^n\{\widetilde{V}_i\widetilde{V}_i^{\mathrm{T}}-\Omega(T_i)\}+\frac{n}{2}\Sigma_\lambda(\beta^{(0)})\right)\beta=\sum_{i=1}^n\widetilde{V}_iY_i. \tag{6.3.5}$$

从而可以得到如下迭代计算步骤.

步骤 1　给定初始值 $\beta^{(0)}$.

步骤 2　令 $\beta^{(0)}=\beta^{(k)}$, 通过方程 (6.3.5) 求解 $\beta^{(k+1)}$.

步骤 3　重复计算步骤 2 直到收敛, 并记 β 的最终估计为 $\hat{\beta}$.

在步骤 1 中, 可以基于式 (6.3.3) 右边的前两项, 并利用普通最小二乘方法给出 β 的初始解. 为实施该算法, 内节点个数 K 和惩罚函数中的截断参数 a 和 λ 需要选择. Fan 和 Li (2001) 表明模型的变量选择结果对 a 的选取不是太敏感, 并建议取 $a=3.7$. 因此在本节中也采用 $a=3.7$. 另外, 类似于 Wang 等 (2008), 用 "交错核实法" 选取 K 和 λ. 具体地, 可以通过最小化如下交错核实目标函数来估计 K 和 λ.

$$\mathrm{CV}(K,\lambda)=\sum_{i=1}^n\left\{Y_i-\widetilde{V}_i^{\mathrm{T}}\hat{\beta}_{[i]}\right\}^2-\sum_{i=1}^n\hat{\beta}_{[i]}^{\mathrm{T}}\Omega(T_i)\hat{\beta}_{[i]}, \tag{6.3.6}$$

其中 $\hat{\beta}_{[i]}$ 为去掉第 i 次观测后, 式 (6.3.3) 的解.

Zhao 和 Xue (2011) 在一定正则条件下证明了所得系数函数估计的相合性和稀疏性. 对于定理的详细讨论和证明, 读者可参考 Zhao 和 Xue (2011).

6.4 部分线性变系数测量误差模型的经验似然统计推断

作为变系数模型和部分线性模型的推广, 模型 (6.1.2) 近年来得到了广泛的关注. 但是已有研究大都是在协变量能够直接观测的情形下进行的讨论. 在实际操作中, 协变量 X_i, Z_i 往往不能直接观测到或不能被准确地测量, 所能观测到的是受了误差干扰的变量 V_i 和 W_i, 即

$$V_i = X_i + e_i, \tag{6.4.1}$$

$$W_i = Z_i + u_i. \tag{6.4.2}$$

对参数部分协变量 X_i 带有测量误差的部分线性变系数模型, You 和 Chen (2006) 在测量误差向量协方差阵已知的情形下, 获得了模型中参数和非参数分量的估计, 并证明了估计量的相合性和渐近正态性. Wang 等 (2011) 运用经验似然方法构造了模型中参数的经验似然置信域及其极大经验似然估计, 并证明了估计量的渐近正态性. 对于协变量 Z_i 带有测量误差的部分线性变系数模型, 冯三营等 (2011) 运用经验似然方法研究了模型中参数和非参数分量的估计问题. 本节分别介绍 Wang 等 (2011) 和冯三营等 (2011) 的估计方法和结论.

6.4.1 参数部分协变量带有测量误差的情形

考虑如下部分线性变系数 EV 模型:

$$\begin{cases} Y_i = X_i^{\mathrm{T}}\beta + Z_i^{\mathrm{T}}\alpha(T_i) + \varepsilon_i, \\ V_i = X_i + e_i, \end{cases} \quad i = 1, \cdots, n, \tag{6.4.3}$$

其中模型误差 ε_i 独立同分布且有 $E(\varepsilon_i|X_i, Z_i, T_i) = 0$. 测量误差 e_i 与 $(X_i, Z_i, T_i, \varepsilon_i)$ 独立且其均值为 0 协方差阵为 Σ_{ee}. 考虑到模型的可识别性, 这里假定 Σ_{ee} 已知.

假设 β 已知, 则模型 (6.4.3) 中的第一个式子可以写成

$$Y_i - \sum_{j=1}^{p} X_{ij}\beta_j = \sum_{j=1}^{q} Z_{ij}\alpha_j(T_i) + \varepsilon_i, \quad i = 1, \cdots, n, \tag{6.4.4}$$

其中 X_{ij} 和 Z_{ij} 分别是 X_i 和 Z_i 的第 j 个元素. 模型 (6.4.4) 可看成是一个变系数模型, 首先利用局部线性方法估计未知系数函数 $\{\alpha_j(\cdot), j = 1, \cdots, q\}$. 当 T_i 在 T 的一个小邻域内时, 可以用一个线性函数局部地逼近 $\alpha_j(T_i)$:

$$\alpha_j(T_i) \approx \alpha_j(T) + \alpha_j'(T)(T_i - T) =: a_j + b_j(T_i - T), \quad j = 1, \cdots, q, \tag{6.4.5}$$

其中 $\alpha_j'(T) = \partial\alpha_j(T)/\partial T$. 注意到,

$$E[Z_i^{\mathrm{T}}|T_i]\alpha(T_i) = E[Y_i - X_i^{\mathrm{T}}\beta|T_i] = E[Y_i - V_i^{\mathrm{T}}\beta|T_i].$$

因此可以通过极小化下面的加权最小二乘问题, 求 a_j 和 b_j:

$$\sum_{i=1}^{n}\left\{\left(Y_i - \sum_{j=1}^{p} V_{ij}\beta_j\right) - \sum_{j=1}^{q}[a_j + b_j(T_i - T)]Z_{ij}\right\}^2 K_h(T_i - T), \tag{6.4.6}$$

其中 $K(\cdot)$ 为核函数, h 为窗宽, 且 $K_h(\cdot) = K(\cdot/h)/h$.

记 $Y = (Y_1, \cdots, Y_n)^{\mathrm{T}}$, $\varepsilon = (\varepsilon_1, \cdots, \varepsilon_n)^{\mathrm{T}}$, $\omega_T = \mathrm{diag}(K_h(T_1 - T), \cdots, K_h(T_n - T))$, $X = (X_1, \cdots, X_n)^{\mathrm{T}}$, $V = (V_1, \cdots, V_n)^{\mathrm{T}}$,

$$D_T = \begin{pmatrix} Z_1^{\mathrm{T}} & \dfrac{T_1 - T}{h} Z_1^{\mathrm{T}} \\ \vdots & \vdots \\ Z_n^{\mathrm{T}} & \dfrac{T_n - T}{h} Z_n^{\mathrm{T}} \end{pmatrix}, \quad M = \begin{pmatrix} Z_1^{\mathrm{T}}\alpha(T_1) \\ \vdots \\ Z_n^{\mathrm{T}}\alpha(T_n) \end{pmatrix}.$$

则式 (6.4.6) 的解为

$$(\hat{a}_1, \cdots, \hat{a}_q, h\hat{b}_1, \cdots, h\hat{b}_q)^{\mathrm{T}} = (D_T^{\mathrm{T}}\omega_T D_T)^{-1} D_T^{\mathrm{T}}\omega_T(Y - V\beta). \tag{6.4.7}$$

从而可得系数函数 $\{\alpha_j(\cdot), j = 1, \cdots, q\}$ 的伪估计为

$$\hat{\alpha}(T) = (\hat{\alpha}_1(T), \cdots, \hat{\alpha}_q(T))^{\mathrm{T}} = (I_q\ 0_q)(D_T^{\mathrm{T}}\omega_T D_T)^{-1} D_T^{\mathrm{T}}\omega_T(Y - V\beta). \tag{6.4.8}$$

M 的估计为

$$\hat{M} = \begin{pmatrix} (Z_1^{\mathrm{T}}\ 0_q^{\mathrm{T}})(D_{T_1}^{\mathrm{T}}\omega_{T_1} D_{T_1})^{-1} D_{T_1}^{\mathrm{T}}\omega_{T_1} \\ \vdots \\ (Z_n^{\mathrm{T}}\ 0_q^{\mathrm{T}})(D_{T_n}^{\mathrm{T}}\omega_{T_n} D_{T_n})^{-1} D_{T_n}^{\mathrm{T}}\omega_{T_n} \end{pmatrix}(Y - V\beta) =: S(Y - V\beta), \tag{6.4.9}$$

其中

$$S = \begin{pmatrix} (Z_1^{\mathrm{T}}\ 0_q^{\mathrm{T}})(D_{T_1}^{\mathrm{T}}\omega_{T_1} D_{T_1})^{-1} D_{T_1}^{\mathrm{T}}\omega_{T_1} \\ \vdots \\ (Z_n^{\mathrm{T}}\ 0_q^{\mathrm{T}})(D_{T_n}^{\mathrm{T}}\omega_{T_n} D_{T_n})^{-1} D_{T_n}^{\mathrm{T}}\omega_{T_n} \end{pmatrix}.$$

令

$$\widetilde{Y}_i = Y_i - \sum_{k=1}^n S_{ik}Y_k, \quad \widetilde{X}_{ij} = X_{ij} - \sum_{k=1}^n S_{ik}X_{kj}, \quad \widetilde{V}_{ij} = V_{ij} - \sum_{k=1}^n S_{ik}V_{kj},$$

其中 S_{ik} 为矩阵 S 的第 (i,k) 个元素. 于是, 引入辅助随机向量

$$\xi_i(\beta) = \widetilde{V}_i(Y_i - V_i^{\mathrm{T}}\beta - Z_i^{\mathrm{T}}\hat{\alpha}(T_i)) + \Sigma_{ee}\beta = \widetilde{V}_i(\widetilde{Y}_i - \widetilde{V}_i^{\mathrm{T}}\beta) + \Sigma_{ee}\beta. \tag{6.4.10}$$

注意在式 (6.4.10) 的右边减了一项 $\Sigma_{ee}\beta$, 目的是避免由于测量误差存在造成的对参数分量的过低估计. 因此, 可以定义 β 的经验似然比函数

$$l(\beta) = \max\left\{\prod_{i=1}^n (np_i) \,\middle|\, p_i \geqslant 0,\ \sum_{i=1}^n p_i = 1,\ \sum_{i=1}^n p_i\xi_i(\beta) = 0\right\}. \tag{6.4.11}$$

由 Lagrange 乘子法, 简单计算可得

$$\log(l(\beta)) = -\sum_{i=1}^n \log\{1 + \lambda^{\mathrm{T}}\xi_i(\beta)\}, \tag{6.4.12}$$

其中 λ 是 Lagrange 乘子, 满足

$$\frac{1}{n}\sum_{i=1}^n \frac{\xi_i(\beta)}{1 + \lambda^{\mathrm{T}}\xi_i(\beta)} = 0. \tag{6.4.13}$$

极大化 $\log(l(\beta))$ 可得 β 的极大经验似然估计 (MELE)$\widetilde{\beta}$. 若 $\log(l(\beta))$ 在 $\hat{\beta}$ 点取得最大值, 且有 $p_i(\hat{\beta}) = 1/n(1 \leqslant i \leqslant n)$, 则

$$\hat{\beta} = \left(\sum_{i=1}^n \widetilde{V}_i\widetilde{V}_i^{\mathrm{T}} - n\Sigma_{ee}\right)^{-1}\sum_{i=1}^n \widetilde{V}_i\widetilde{Y}_i. \tag{6.4.14}$$

由 $\widetilde{\beta}$ 及式 (6.4.8) 可以定义系数函数 $\{\alpha_j(\cdot), j = 1, \cdots, q\}$ 的估计为

$$\widetilde{\alpha}(T) = (\widetilde{\alpha}_1(T), \cdots, \widetilde{\alpha}_q(T))^{\mathrm{T}} = (I_q\ 0_q)(D_T^{\mathrm{T}}\omega_T D_T)^{-1}D_T^{\mathrm{T}}\omega_T(Y - V\widetilde{\beta}). \tag{6.4.15}$$

为了得到本节的主要结论, 需要给出一些正则条件.

(C6.4.1) 随机变量 T 有有界支撑 $\mathcal{T}$, 其密度函数 $f(\cdot)$ 是 Lipschitz 连续的, 且 $f(\cdot) > 0$.

(C6.4.2) $\{\alpha_j(\cdot), j = 1, \cdots, q\}$ 在 $\mathcal{T}$ 上有二阶连续导数.

(C6.4.3) $K(\cdot)$ 为对称概率密度核函数且具有紧支撑. 当 $n \to \infty$ 时, $nh^2/\log^2 n \to \infty$, $nh^8 \to 0$.

(C6.4.4) 矩阵 $\Gamma(T) = E(Z_1Z_1^{\mathrm{T}} \mid T)$ 非退化. $E(X_1X_1^{\mathrm{T}} \mid T), \Gamma^{-1}(T)$ 和 $\Phi(T) =: E(Z_1X_1^{\mathrm{T}} \mid T)$ 均为 Lipschitz 连续.

(C6.4.5) 存在常数 $s > 2$ 使得 $E\|\varepsilon_1\|^{2s} < \infty$, $E\|e_1\|^{2s} < \infty$, $E\|X_1\|^{2s} < \infty$, $E\|Z_1\|^{2s} < \infty$. 对某个 $\delta < 2 - s^{-1}$, 当 $n \to \infty$ 时, 有 $n^{2\delta-1}h \to \infty$.

定理 6.4.1　假设条件(C6.4.1)~条件(C6.4.5)成立, 如果 β_0 是参数真值, 当 $n \to \infty$ 时, 则有

$$-2\log(l(\beta_0)) \xrightarrow{L} \chi_p^2.$$

基于定理 6.4.1, 对任给的 $0<\alpha<1$, 存在 c_α 使得 $P(\chi_p^2 > c_\alpha)=\alpha$, 则

$$I_\alpha(\beta^*) = \left\{\beta^* \in \mathbb{R}^p \;\middle|\; -2\log(l(\beta^*)) \leqslant c_\alpha\right\}$$

为参数向量 β 的具有渐近置信水平 $1-\alpha$ 的置信域.

定理 6.4.2　假设条件(C6.4.1)~条件(C6.4.5)成立, 当 $n \to \infty$ 时, 则有

$$\sqrt{n}(\widetilde{\beta}-\beta) \xrightarrow{L} N\left(0, \Sigma_1^{-1}\Lambda_1\Sigma_1^{-1}\right),$$

其中

$$\Lambda_1 = E\{(\varepsilon - e^{\mathrm{T}}\beta)(X - \Phi^{\mathrm{T}}(T)\Gamma^{-1}(T)Z)\}^{\otimes 2} + E\left\{\left(ee^{\mathrm{T}} - \Sigma_{ee}\right)\beta\right\}^{\otimes 2} + E(ee^{\mathrm{T}}\varepsilon^2),$$

$$\Sigma_1 = E(XX^{\mathrm{T}}) - E(\Phi^{\mathrm{T}}(T)\Gamma^{-1}(T)\Phi(T)).$$

如果 ε 具有同方差 σ^2 并独立于 (X, Z, e, T), 则

$$\Lambda_1 = E(\varepsilon - e^{\mathrm{T}}\beta)^2\Sigma_1 + E\left\{\left(ee^{\mathrm{T}} - \Sigma_{ee}\right)\beta\right\}^{\otimes 2} + \Sigma_{ee}\sigma^2.$$

6.4.2　非参数部分协变量带有测量误差的情形

考虑如下部分线性变系数 EV 模型:

$$\begin{cases} Y_i = X_i^{\mathrm{T}}\beta + Z_i^{\mathrm{T}}\alpha(T_i) + \varepsilon_i, \\ W_i = Z_i + u_i, \end{cases} \qquad i = 1, \cdots, n, \tag{6.4.16}$$

其中 ε_i 是相互独立的模型误差, 且 $E(\varepsilon_i)=0$, $\mathrm{Var}(\varepsilon_i)=\sigma^2<\infty$, $\quad i=1,2,\cdots,n$. 测量误差 $\{u_i\}_{i=1}^n$ 独立同分布, 其均值为 0, 协方差为 Σ_{uu}, 且 u_i 与 $(X_i, Z_i, T_i, \varepsilon_i)$ 相互独立. 考虑到模型的可识别性, 这里假定 Σ_{uu} 已知.

若 Z_i 精确观测, 即无测量误差, 则由系数函数 $\{\alpha_j(\cdot), j=1,\cdots,q\}$ 的局部线性逼近式 (6.4.5), 极小化下面的加权最小二乘问题

$$\sum_{i=1}^n \left\{\left(Y_i - \sum_{j=1}^p X_{ij}\beta_j\right) - \sum_{j=1}^q [a_j + b_j(T_i - T)]Z_{ij}\right\}^2 K_h(T_i - T), \tag{6.4.17}$$

可得 a_j 和 b_j 的估计为

$$(\hat{a}_1, \cdots, \hat{a}_q, h\hat{b}_1, \cdots, h\hat{b}_q)^{\mathrm{T}} = \{D_T^{\mathrm{T}}\omega_T D_T\}^{-1}D_T^{\mathrm{T}}\omega_T(Y - X\beta). \tag{6.4.18}$$

现在, 由于 Z_i 不可观测, 观测到的是带有误差扰动的 W_i, 如果在式 (6.4.18) 中直接用 W_i 替代 Z_i, 则该估计不再是相合估计. 为了消除由测量误差所带来的估计偏差, 借鉴 You 等 (2006) 的思想, 对式 (6.4.18) 进行如下的局部纠偏:

$$(\hat{a}_1,\cdots,\hat{a}_q,h\hat{b}_1,\cdots,h\hat{b}_q)^{\mathrm{T}}=\{(D_T^W)^{\mathrm{T}}\omega_T D_T^W-\Omega\}^{-1}(D_T^W)^{\mathrm{T}}\omega_T(Y-X\beta), \quad (6.4.19)$$

于是系数函数 $\{\alpha_j(\cdot),j=1,\cdots,q\}$ 的估计为

$$\begin{aligned}\hat{\alpha}(T)&=(\hat{\alpha}_1(T),\cdots,\hat{\alpha}_q(T))^{\mathrm{T}}\\&=(I_q\ 0_q)\{(D_T^W)^{\mathrm{T}}\omega_T D_T^W-\Omega\}^{-1}(D_T^W)^{\mathrm{T}}\omega_T(Y-X\beta),\end{aligned} \quad (6.4.20)$$

其中 D_T^W 是由 W_i 替换 D_T 中的 Z_i 所得, 且

$$\Omega=\sum_{i=1}^n\Sigma_{uu}\otimes\begin{pmatrix}1 & \dfrac{T_i-T}{h}\\ \dfrac{T_i-T}{h} & \left(\dfrac{T_i-T}{h}\right)^2\end{pmatrix}K_h(T_i-T).$$

记

$$\widetilde{S}=\begin{pmatrix}(W_1^{\mathrm{T}}\ 0_q^{\mathrm{T}})\{(D_{T_1}^W)^{\mathrm{T}}\omega_{T_1}D_{T_1}^W-\Omega\}^{-1}(D_{T_1}^W)^{\mathrm{T}}\omega_{T_1}\\ \vdots\\ (W_n^{\mathrm{T}}\ 0_q^{\mathrm{T}})\{(D_{T_n}^W)^{\mathrm{T}}\omega_{T_n}D_{T_n}^W-\Omega\}^{-1}(D_{T_n}^W)^{\mathrm{T}}\omega_{T_n}\end{pmatrix},$$

$$Q_i=(I_q\ 0_q)\{(D_{T_i}^W)^{\mathrm{T}}\omega_{T_i}D_{T_i}^W-\Omega\}^{-1}(D_{T_i}^W)^{\mathrm{T}}\omega_{T_i},$$

$$\widetilde{Y}_i=Y_i-\sum_{k=1}^n\widetilde{S}_{ik}Y_k,\quad \widetilde{X}_{ij}=X_{ij}-\sum_{k=1}^n\widetilde{S}_{ik}X_{kj},$$

其中 $\widetilde{S}_{ik}$ 为 $\widetilde{S}$ 的第 (i,k) 元素, 则 β 的 profile 最小二乘估计为

$$\begin{aligned}\hat{\beta}&=\arg\min_{\beta\in\mathbb{R}^p}\sum_{i=1}^n\left[\{Y_i-X_i^{\mathrm{T}}\beta-W_i^{\mathrm{T}}\hat{\alpha}(T_i)\}^2-\hat{\alpha}^{\mathrm{T}}(T_i)\Sigma_{uu}\hat{\alpha}(T_i)\right],\\&=\arg\min_{\beta\in\mathbb{R}^p}\sum_{i=1}^n\left[\{\widetilde{Y}_i-\widetilde{X}_i^{\mathrm{T}}\beta\}^2-\hat{\alpha}^{\mathrm{T}}(T_i)\Sigma_{uu}\hat{\alpha}(T_i)\right].\end{aligned} \quad (6.4.21)$$

上式右端第二项的作用是修正由测量误差引起的估计偏差. 注意到 $\hat{\alpha}(T_i)=Q_i(Y-X\beta)$, 则上式右端对 β 求偏导并令其为零, 可得估计方程

$$\sum_{i=1}^n\left[\widetilde{X}_i(\widetilde{Y}_i-\widetilde{X}_i^{\mathrm{T}}\beta)-X^{\mathrm{T}}Q_i^{\mathrm{T}}\Sigma_{uu}Q_i(Y-X\beta)\right]=0.$$

记

$$\eta_i(\beta) = \widetilde{X}_i(\widetilde{Y}_i - \widetilde{X}_i^{\mathrm{T}}\beta) - X^{\mathrm{T}}Q_i^{\mathrm{T}}\Sigma_{uu}Q_i(Y - X\beta). \tag{6.4.22}$$

于是由 Owen (1991), 可以定义经验对数似然比函数

$$\mathcal{R}(\beta) = -2\max\left\{\sum_{i=1}^{n}\log(np_i)\,\middle|\, p_i \geqslant 0,\ \sum_{i=1}^{n}p_i = 1,\ \sum_{i=1}^{n}p_i\eta_i(\beta) = 0\right\}. \tag{6.4.23}$$

由 Lagrange 乘子法, $\mathcal{R}(\beta)$ 可以表示为

$$\mathcal{R}(\beta) = 2\sum_{i=1}^{n}\log\{1 + \lambda^{\mathrm{T}}\eta_i(\beta)\}, \tag{6.4.24}$$

其中 λ 是 Lagrange 乘子, 满足

$$\frac{1}{n}\sum_{i=1}^{n}\frac{\eta_i(\beta)}{1 + \lambda^{\mathrm{T}}\eta_i(\beta)} = 0. \tag{6.4.25}$$

极小化 $\mathcal{R}(\beta)$ 可得 β 的极大经验似然估计 $\widetilde{\beta}$. 由 $\widetilde{\beta}$ 及式 (6.4.20) 可以定义系数函数 $\{\alpha_j(\cdot), j = 1, \cdots, q\}$ 的估计为

$$\begin{aligned}\widetilde{\alpha}(T) &= (\widetilde{\alpha}_1(T), \cdots, \widetilde{\alpha}_q(T))^{\mathrm{T}}\\ &= (I_q\ 0_q)\{(D_T^W)^{\mathrm{T}}\omega_T D_T^W - \Omega\}^{-1}(D_T^W)^{\mathrm{T}}\omega_T(Y - X\widetilde{\beta}).\end{aligned} \tag{6.4.26}$$

为了得到本节的主要结果, 需要下面的条件:

(C6.4.5$'$) 存在常数 $s > 2$ 使得 $E\|\varepsilon_1\|^{2s} < \infty$, $E\|u_1\|^{2s} < \infty$, $E\|X_1\|^{2s} < \infty$, $E\|Z_1\|^{2s} < \infty$. 对某个 $\delta < 2 - s^{-1}$, 当 $n \to \infty$ 时, 有 $n^{2\delta-1}h \to \infty$.

定理 6.4.3　假设条件(C6.4.1)~条件(C6.4.4)和条件(C6.4.5$'$)成立, 如果 β_0 是参数真值, 当 $n \to \infty$ 时, 则有

$$\mathcal{R}(\beta_0) \xrightarrow{L} \chi_p^2.$$

基于定理 6.4.3, 可以定义参数向量 β 的置信域, 即对任给的 $0 < \alpha < 1$, 存在 c_α 使得 $P\{\chi_p^2 \leqslant c_\alpha\} = 1 - \alpha$, 则

$$I_\alpha(\beta^*) = \left\{\beta^* \in \mathbb{R}^p \,\middle|\, \mathcal{R}(\beta^*) \leqslant c_\alpha\right\}$$

为参数向量 β 的具有渐近置信水平 $1 - \alpha$ 的置信域, 并且有 $P(\beta \in I_\alpha(\beta^*)) = 1 - \alpha + o(1)$.

定理 6.4.4 假设条件(C6.4.1)~条件(C6.4.4)和条件(C6.4.5′)成立, 如果 β_0 是参数真值, 当 $n\to\infty$ 时, 则有

$$\sqrt{n}(\widetilde{\beta}-\beta_0)\xrightarrow{L}N\left(0,\Sigma_2^{-1}\Lambda_2\Sigma_2^{-1}\right),$$

其中 $\Sigma_2=E(X_1X_1^{\mathrm{T}})-E(\Phi^{\mathrm{T}}(T_1)\Gamma^{-1}(T_1)\Phi(T_1))$,

$$\begin{aligned}\Lambda_2=&E(\varepsilon_1-u_1^{\mathrm{T}}\alpha(T_1))^2\Sigma_1+\sigma^2E\left\{\Phi^{\mathrm{T}}(T_1)\Gamma^{-1}(T_1)\Sigma_{uu}\Gamma^{-1}(T_1)\Phi(T_1)\right\}\\&+E\left\{\Phi^{\mathrm{T}}(T_1)\Gamma^{-1}(T_1)\left(u_1u_1^{\mathrm{T}}-\Sigma_{uu}\right)\alpha(T_1)\right\}^{\otimes 2}.\end{aligned}$$

定理 6.4.5 假设条件(C6.4.1)~条件(C6.4.4)和条件(C6.4.5′)成立, 如果 $\alpha(T)$ 是系数函数真值, 当 $n\to\infty$ 时, 则有

$$\sqrt{nh}\left(\widetilde{\alpha}(T)-\alpha(T)-\frac{1}{2}h^2\frac{\mu_2^2-\mu_1\mu_3}{\mu_2-\mu_1^2}\alpha''(T)\right)\xrightarrow{L}N(0,\Delta),$$

其中

$$\begin{aligned}\Delta&=(\kappa_0^2\nu_0+2\kappa_0\kappa_1\nu_1+\kappa_1^2\nu_2)f(T)^{-1}\Sigma^*,\\\kappa_0&=\mu_2/(\mu_2-\mu_1^2),\\\kappa_1&=-\mu_1/(\mu_2-\mu_1^2),\\\Sigma^*&=\Gamma^{-1}(T)[\sigma^2\Gamma(T)+\sigma^2\Sigma_{uu}\\&\quad+E\{\zeta_1\alpha(T)\alpha^{\mathrm{T}}(T)\zeta_1^{\mathrm{T}}|T\}]\Gamma^{-1}(T),\\\zeta_1&=\Sigma_{uu}-u_1u_1^{\mathrm{T}}-X_1u_1^{\mathrm{T}}.\end{aligned}$$

6.4.3 模拟研究

例 6.4.1 考虑模型:

$$Y=X^{\mathrm{T}}\beta+Z^{\mathrm{T}}\alpha(T)+\varepsilon,\quad V=X+e,$$

其中 $\beta=1$, 系数函数 $\alpha(t)=(\alpha_1(t),\alpha_2(t))^{\mathrm{T}}$, $\alpha_1(t)=4+\sin(2\pi t)$, $\alpha_2(t)=2t(1-t)$. 协变量 $X\sim N(1,1)$, $Z_1\sim N(1,1)$, $Z_2\sim N(1,1)$, T 服从区间 [0,1] 上均匀分布, 模型误差 $\varepsilon\sim N(0,0.4^2)$, 测量误差 $e\sim N(0,\sigma_e^2)$.

在下面的模拟中, 分别取样本容量 n=100, 150, 300, 测量误差方差 $\sigma_e^2=0.2^2$, 0.4^2, 0.6^2. 取核函数为 Epanechnikov 核 $K(t)=0.75(1-t^2)I(|t|\leqslant 1)$, 窗宽 h 采用实际中广泛应用的交错核实方法进行选择. 定义窗宽 h 的交错核实得分为

$$\mathrm{CV}(h)=\frac{1}{n}\sum_{i=1}^{n}\{Y_i-V_i^{\mathrm{T}}\widetilde{\beta}_{-i}-Z_i^{\mathrm{T}}\widetilde{\alpha}_{h,-i}(T_i)\}^2,$$

其中 $\widetilde{\beta}_{-i}$ 是去掉第 i 个个体后, 极大化 $\log(l(\beta))$ 得到的极大经验似然估计, $\widetilde{\alpha}_{h,-i}(T_i)$ 为相应的由式 (6.4.15) 得到的系数函数的估计. 从而光滑参数 $h_{\rm cv} = \arg\min\limits_h {\rm CV}(h)$. 在本例中, 考虑采用经验似然 (EL) 和最小二乘 (LS) 两种方法构造未知参数的置信域. 所采用的最小二乘方法是基于 You 和 Chen (2006) 所给出的估计的渐近正态分布. 重复模拟 1000 次, 比较两种方法得到的参数 β 的置信水平为 95% 的置信域的覆盖概率 (CP) 和平均区间长度 (AL).

从表 6.4.1 中的模拟结果可以看出:

(1) 在三种不同的测量误差下, 随着样本量 n 的增加, 由经验似然方法和最小二乘方法得到的置信区间的覆盖概率都趋于 0.95, 相应置信区间的平均长度都减少;

(2) 就置信区间的精度而言, 经验似然方法要优于最小二乘方法, 因为从表 6.4.1 中容易发现基于经验似然方法得到的置信区间的平均长度比最小二乘法短且覆盖概率高;

(3) 在同样的样本容量下, 表 6.4.1 还给出了在三种不同测量误差 $\sigma_e^2 = 0.2^2$, 0.4^2, 0.6^2 下模拟结果的比较. 当测量误差 $\sigma_e^2 = 0.2^2$ 时, 经验似然和最小二乘表现最好. 而且, 随着测量误差 σ_e^2 的增大, 置信区间的覆盖概率减小, 而平均长度增大.

表 6.4.1　β 的置信水平为95%的置信区间的覆盖概率和平均区间长度

σ_e^2	n	CP		AL	
		EL	LS	EL	LS
0.2^2	100	0.9420	0.9280	0.2524	0.2675
	150	0.9530	0.9280	0.1813	0.1820
	300	0.9560	0.9350	0.1437	0.1458
0.4^2	100	0.9360	0.9110	0.2981	0.3426
	150	0.9310	0.9150	0.2153	0.2344
	300	0.9430	0.9270	0.1728	0.1794
0.6^2	100	09120	0.8640	0.3721	0.4159
	150	0.9160	0.8740	0.2962	0.3478
	300	0.9280	0.8850	0.2401	0.3114

例 6.4.2　考虑模型:

$$Y = X^{\rm T}\beta + Z\alpha(T) + \varepsilon, \quad W = Z + u,$$

其中 $\beta = (1,2)^{\rm T}$, 系数函数 $\alpha(t) = 2t(1-t)$, $(X^{\rm T}, Z)^{\rm T}$ 服从多元正态分布, 其均值向量为 $(1,1,1)^{\rm T}$, 协方差阵 $\Sigma = (\sigma_{ij})$ 满足 $\sigma_{ij} = 0.5^{|i-j|}$. T 服从区间 [0,1] 上均匀分布, 模型误差 $\varepsilon \sim N(0,1)$, 测量误差 u 服从均值为 0 方差为 σ_u^2 的正态分布.

在下面的模拟中, 分别取样本容量 n=30, 100, 200, 400, 测量误差方差 $\sigma_u^2 = 0.4^2$, 0.8^2. 取核函数为 Epanechnikov 核 $K(t) = 0.75(1-t^2)I(|t| \leqslant 1)$, 窗宽采用 CV

方法进行选择, 重复模拟 2000 次. 模拟结果见表 6.4.2, 图 6.4.1 和图 6.4.2, 其中图 6.4.1 给出的是测量误差方差 $\sigma_u^2 = 0.4^2$ 时 β 在不同样本容量下置信水平为 0.95 的置信域的模拟结果, 图 6.4.2 给出的是系数函数 $\alpha(t)$ 在 $n = 200$, $\sigma_u^2 = 0.4^2$ 时的模拟结果, 其他情形下的模拟效果图与图 6.4.1 和图 6.4.2 类似.

表 6.4.2 β 的置信水平为95%的置信域的覆盖概率

n	$\sigma_u^2 = 0.4^2$	$\sigma_u^2 = 0.8^2$
30	0.8930	0.8840
100	0.9320	0.9260
200	0.9420	0.9350
400	0.9480	0.9440

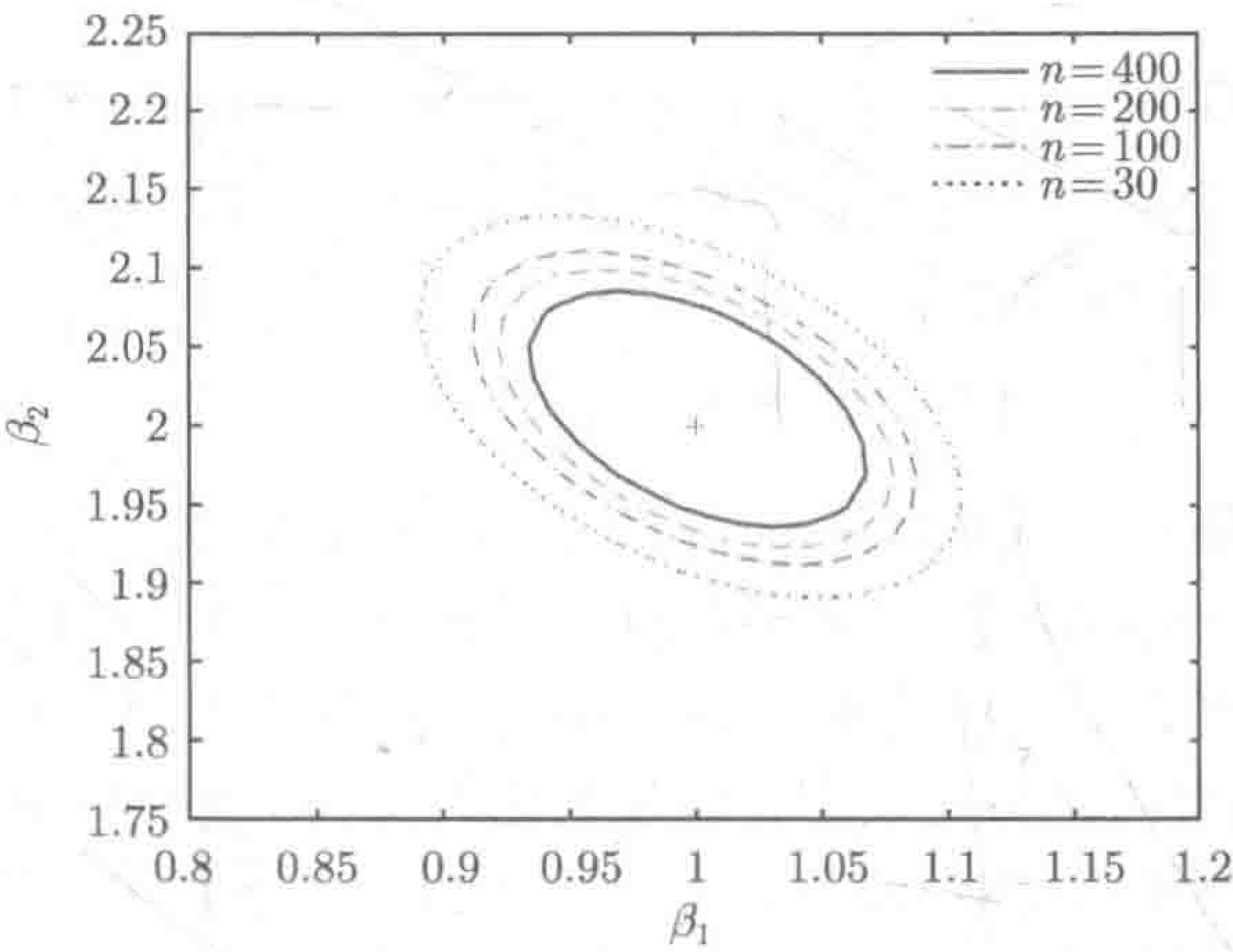


图 6.4.1 (β_1, β_2) 的置信水平为 0.95 的置信域

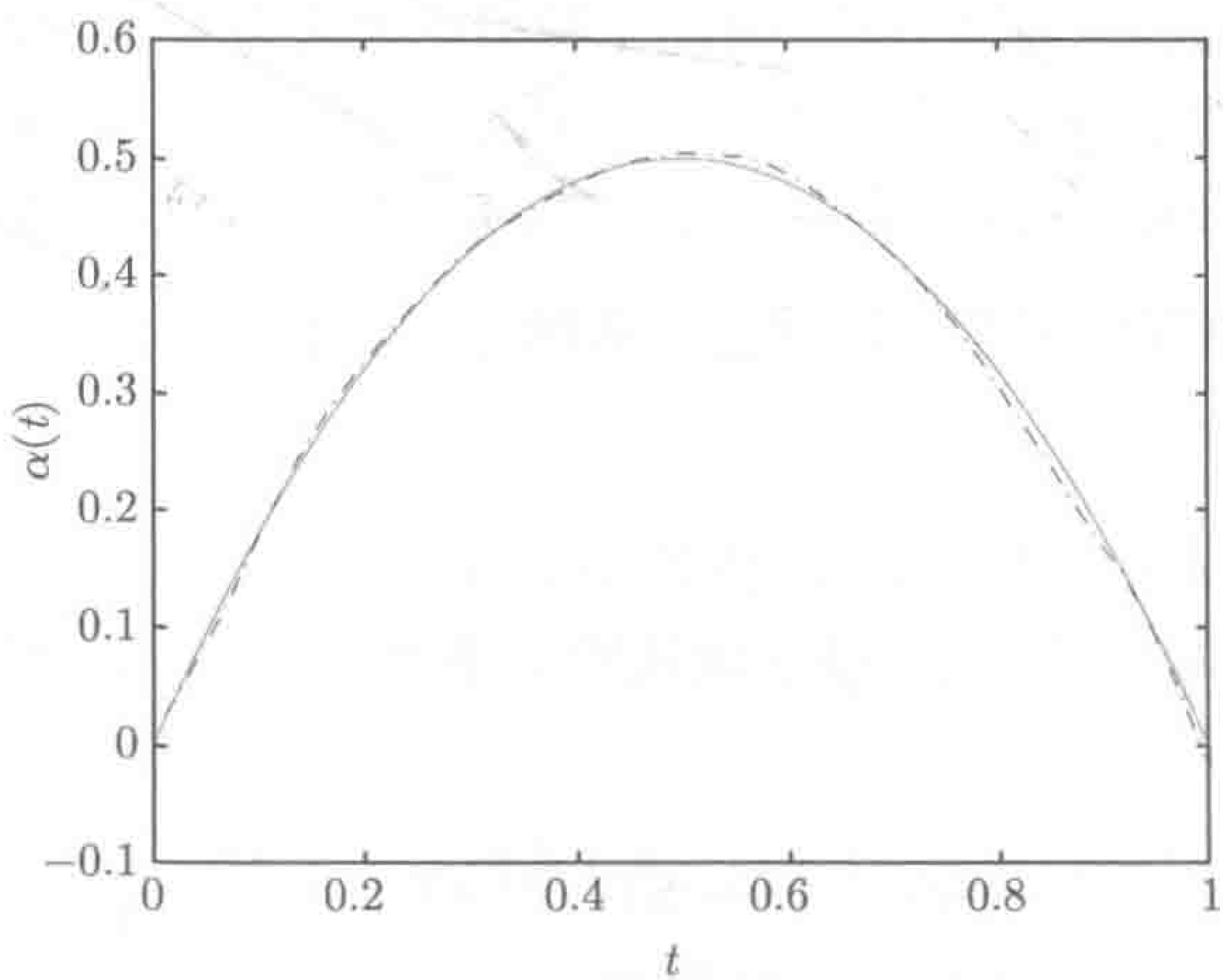


图 6.4.2 $\alpha(t)$ 的估计曲线 (虚线) 和真实曲线 (实线)

从表 6.4.2, 图 6.4.1 和图 6.4.2 所得到的模拟结果可以看出:

(1) 测量误差方差 σ_u^2 的大小影响置信域的精度, σ_u^2 增大时, 参数 β 的覆盖概率逐渐变小;

(2) 随着样本容量 n 的增加, β 的覆盖概率变大并趋近于置信水平 0.95, 且置信域逐渐变小;

(3) 对比图 6.4.2 中系数函数 $\alpha(t)$ 的估计曲线和真实曲线, 可以看出所提出的局部纠偏估计方法是理想的.

6.5 部分线性变系数测量误差模型的约束统计推断

6.5.1 引言

在许多重要的统计应用中, 除了样本信息以外, 人们基于历史研究等原因往往还会获得一些关于回归参数的先验信息, 这些先验信息可以有助于提高参数估计的精度. 这里考虑如下约束条件

$$H\beta = d, \tag{6.5.1}$$

其中 H 为 $k \times p$ 维已知行满秩矩阵, d 为 $k \times 1$ 已知向量. 对于模型 (6.1.2), Fan 和 Huang (2005) 分别构造了 GLR 检验统计量和 Wald 检验统计量来检验 $H\beta = 0$ 问题; 魏传华和吴喜之 (2008) 在约束条件 (6.5.1) 下构造了 profile Lagrange 算子检验. 上述文献都没有考虑协变量含有测量误差的情况. 当模型 (6.1.2) 中的参数部分协变量 X 带有测量误差时, Zhang 等 (2011) 提出了参数和非参数分量的约束调整 profile 最小二乘估计, 并构造了调整的 profile Lagrange 算子检验统计量. 当非参数部分协变量 Z 带有测量误差时, Feng 和 Xue (2014) 结合 profile 最小二乘、局部偏差纠正及 Lagrange 算子方法, 提出了模型中参数分量和非参数系数函数的两种约束估计, 并构造了未知参数向量 β 的局部偏差纠正 profile Lagrange 算子检验统计量. 本节分别介绍 Zhang 等 (2011) 与 Feng 和 Xue (2014) 的估计方法和结论.

6.5.2 参数部分协变量带有测量误差的情形

1. 估计方法

本节仍采用 6.4.1 节中的记号. 假设 $\{(Y_i; V_i^{\mathrm{T}}, Z_i^{\mathrm{T}}, T_i), 1 \leqslant i \leqslant n\}$ 是来自模型 (6.4.3) 的独立同分布随机样本, 基于线性约束条件 (6.5.1), β 的修正的辅助函数定义为

$$\begin{aligned} F(\beta, \lambda) = &[(I-S)Y - (I-S)V\beta]^{\mathrm{T}}[(I-S)Y - (I-S)V\beta] \\ &- n\beta^{\mathrm{T}}\Sigma_{ee}\beta + 2\lambda^{\mathrm{T}}(H\beta - d). \end{aligned} \tag{6.5.2}$$

计算函数 $F(\beta,\lambda)$ 分别关于 β 和 λ 的偏导数, 并令偏导数等于 0, 有

$$\left\{\begin{array}{l}\dfrac{\partial F(\beta,\lambda)}{\partial\beta}=-2\widetilde{V}^{\mathrm{T}}(I-S)Y+2\widetilde{V}^{\mathrm{T}}\widetilde{V}\beta-2n\Sigma_{ee}\beta+2H^{\mathrm{T}}\lambda=0,\\ \dfrac{\partial F(\beta,\lambda)}{\partial\lambda}=H\beta-d=0.\end{array}\right. \tag{6.5.3}$$

简单计算可得方程 (6.5.3) 的解为

$$\left\{\begin{array}{l}\hat{\beta}^{R}=\hat{\beta}-\left(\widetilde{V}^{\mathrm{T}}\widetilde{V}-n\Sigma_{ee}\right)^{-1}H^{\mathrm{T}}\left[H\left(\widetilde{V}^{\mathrm{T}}\widetilde{V}-n\Sigma_{ee}\right)^{-1}H^{\mathrm{T}}\right]^{-1}(H\hat{\beta}-d),\\ \hat{\lambda}=\left[H\left(\widetilde{V}^{\mathrm{T}}\widetilde{V}-n\Sigma_{ee}\right)^{-1}H^{\mathrm{T}}\right]^{-1}(H\hat{\beta}-d),\end{array}\right. \tag{6.5.4}$$

其中 $\widetilde{V}=(I-S)V$, 且

$$\hat{\beta}=\left[\widetilde{V}^{\mathrm{T}}\widetilde{V}-n\Sigma_{ee}\right]^{-1}\widetilde{V}^{\mathrm{T}}(I-S)Y \tag{6.5.5}$$

是由 You 和 Chen (2006) 提出的修正 profile 最小二乘估计 (MPLSE). $\hat{\beta}^{R}$ 称为参数分量 β 修正的约束 profile 最小二乘估计 (MRPLSE).

由式 (6.4.8) 和式 (6.5.4), 可以定义系数函数 $\alpha(\cdot)$ 的估计为

$$\hat{\alpha}^{R}(T)=(I_q\quad 0_q)\{D_T^{\mathrm{T}}\omega_T D_T\}^{-1}D_T^{\mathrm{T}}\omega_T(Y-V\hat{\beta}^{R}). \tag{6.5.6}$$

2. 渐近性质

定理 6.5.1 假设条件(C6.4.1)~条件(C6.4.5)成立. 当 $n\to\infty$ 时, 则参数分量 β 的修正的约束 profile 最小二乘估计 $\hat{\beta}^{R}$ 是渐近正态的, 即

$$\sqrt{n}(\hat{\beta}^{R}-\beta)\xrightarrow{L}N(0,\Sigma), \tag{6.5.7}$$

其中

$$\begin{aligned}\Sigma=&\Sigma_1^{-1}\Sigma_2\Sigma_1^{-1}-\Sigma_1^{-1}\Sigma_2\Sigma_1^{-1}\Sigma_3\Sigma_1^{-1}-\Sigma_1^{-1}\Sigma_3\Sigma_1^{-1}\Sigma_2\Sigma_1^{-1}\\&+\Sigma_1^{-1}\Sigma_3\Sigma_1^{-1}\Sigma_2\Sigma_1^{-1}\Sigma_3\Sigma_1^{-1},\end{aligned}$$

$$\Sigma_1=E(X_1X_1^{\mathrm{T}})-E\{E(X_1Z_1^{\mathrm{T}}|T_1)E(Z_1Z_1^{\mathrm{T}}|T_1)^{-1}E(X_1Z_1^{\mathrm{T}}|T_1)^{\mathrm{T}}\},$$

$$\Sigma_2=E(\varepsilon_1-e_1^{\mathrm{T}}\beta)^2\Sigma_1+\sigma^2\Sigma_{ee}+E\left\{\left(e_1e_1^{\mathrm{T}}-\Sigma_{ee}\right)\beta\beta^{\mathrm{T}}\left(e_1e_1^{\mathrm{T}}-\Sigma_{ee}\right)\right\},$$

$$\Sigma_3=H^{\mathrm{T}}\left[H\Sigma_1^{-1}H^{\mathrm{T}}\right]^{-1}H.$$

为了应用定理 6.5.1 对参数分量 β 进行统计推断, 首先需要给出渐近协方差矩阵 Σ 的相合估计. 注意到定理 6.5.1 中 Σ 的结构, 仅需要估计 Σ_1 和 Σ_2, 然后使用 plug-in 方法即可得到 Σ 的相合估计. Σ_1 和 Σ_2 的相合估计分别为

$$\hat{\Sigma}_1 = \frac{1}{n}\sum_{i=1}^{n} \widetilde{V}_i \widetilde{V}_i^{\mathrm{T}} - \Sigma_{ee} \tag{6.5.8}$$

和

$$\hat{\Sigma}_2 = \frac{1}{n}\sum_{i=1}^{n} \left\{ \widetilde{V}_i (\widetilde{Y}_i - \widetilde{V}_i^{\mathrm{T}} \hat{\beta}) + \Sigma_{ee}\hat{\beta} \right\}^{\otimes 2}, \tag{6.5.9}$$

其中 $\widetilde{V} = (\widetilde{V}_1, \cdots, \widetilde{V}_n)^{\mathrm{T}}$.

注 6.5.1 定理 6.5.1 表明修正的约束 profile 最小二乘估计 $\hat{\beta}^R$ 是渐近有效的. 当 X 被准确观测时, 容易证得 $\hat{\beta}^R$ 与魏传华和吴喜之 (2008) 得到的估计量有相同的渐近分布.

推论 6.5.1 假设条件(C6.4.1)~条件(C6.4.5)成立. B 是一个 $l \times p$ 矩阵, 当 $n \to \infty$, 则有

$$\sqrt{n}(B\hat{\beta}^R - B\beta) \xrightarrow{L} N\left(0, B\Sigma B^{\mathrm{T}}\right). \tag{6.5.10}$$

注 6.5.2 推论 6.5.1 可用来构造参数分量任意线性组合的置信域, 即对于任意的 $0 < \alpha < 1$, 存在 c_α 使得 $P(\chi_p^2 > c_\alpha) \leqslant 1 - \alpha$, 那么

$$I_\alpha(B\beta^*) = \left\{ B\beta^* \in \mathbb{R}^l \;:\; (B\hat{\beta}^R - B\beta^*)^{\mathrm{T}} \left(\frac{1}{n} B\hat{\Sigma}B^{\mathrm{T}} \right)^{-1} (B\hat{\beta}^R - B\beta^*) \leqslant c_\alpha \right\}$$

构造了线性组合 $B\beta$ 的 $1-\alpha$ 置信域.

定理 6.5.2 假设条件(C6.4.1)~条件(C6.4.5)成立. 当 $n \to \infty$ 时, 则有

$$\max_{1\leqslant j\leqslant q} \sup_{T\in\mathcal{T}} |\hat{\alpha}_j^R(T) - \alpha_j(T)| = O\{h^2 + (\log n/nh)^{\frac{1}{2}}\} \quad \text{a.s.}. \tag{6.5.11}$$

注 6.5.3 定理 6.5.2 提供了 $\alpha(\cdot)$ 估计量 $\hat{\alpha}^R(\cdot)$ 的强收敛速度. 如果取最优窗宽 $h_{\mathrm{opt}} = O(n^{-1/5})$, 则模型 (6.4.3) 中系数函数的估计达到了通常非参数回归的最优强一致收敛速度, 即式 (6.5.11) 成为

$$\max_{1\leqslant j\leqslant q} \sup_{T\in\mathcal{T}} |\hat{\alpha}_j^R(T) - \alpha_j(T)| = O\left\{ n^{-2/5} (\log n)^{1/2} \right\} \quad \text{a.s.}. \tag{6.5.12}$$

定理 6.5.3 假设条件(C6.4.1)~条件(C6.4.5)和 $nh^5 = O(1)$ 成立. 当 $n \to \infty$ 时, 则有

$$\sqrt{nh}\left(\hat{\alpha}^R(T) - \alpha(T) - \frac{h^2}{2}\mu_2 \alpha''(T) \right) \xrightarrow{L} N\Big(0, \Sigma(T)\Big), \tag{6.5.13}$$

其中 $\Sigma(T) = \nu_0 f(T)^{-1} E(\varepsilon_1 - e_1^{\mathrm{T}}\beta)^2 \Gamma(T)^{-1}$, $\Gamma(T) = E(Z_1 Z_1^{\mathrm{T}} \mid T)$, $\nu_0 = \int K^2(t)\mathrm{d}t$.

3. 假设检验

下面考虑线性假设检验问题:

$$H_0: H\beta = d \quad \text{vs} \quad H_1: H\beta = d + n^{-\frac{1}{2}}\delta, \tag{6.5.14}$$

其中 δ 是一个 $k\times 1$ 常数向量. 对于部分线性变系数模型 (6.1.2), Fan 和 Huang (2005) 分别构造了广义似然比 (GLR) 检验统计量和 Wald 检验统计量来检验 $H\beta = 0$ 问题; 魏传华和吴喜之 (2008) 在约束条件 (6.5.1) 下构造了 profile Lagrange 算子检验. 这里考虑参数部分协变量 X_i 不能准确测量, 具有测量误差的情形. 如果忽略测量误差, 直接用可观测变量 V_i 替换变量 X_i, 则所构造的检验统计量的渐近分布很难确定. 为了克服这个困难, 此处提出一个修正的 profile Lagrange 乘子检验方法, 并且在一定的正则条件下, 证明了所提检验统计量在原假设下是渐近 χ^2 分布.

令 $C_0 = [H\hat{\Sigma}_1^{-1}H^{\mathrm{T}}]^{-1}$, $L_0 = H\hat{\Sigma}_1^{-1}\hat{\Sigma}_2\hat{\Sigma}_1^{-1}H^{\mathrm{T}}$, 其中 $\hat{\Sigma}_1^{-1}\hat{\Sigma}_2\hat{\Sigma}_1^{-1}$ 是定理 6.5.1 中定义的 $\Sigma_1^{-1}\Sigma_2\Sigma_1^{-1}$ 的相合估计. 由定义在式 (6.5.4) 中的 Lagrange 乘子估计, 在原假设 H_0 下, 能够构造如下修正的 profile Lagrange 乘子检验统计量

$$\hat{T}_{\mathrm{MPLMS}} = \frac{1}{n}\hat{\lambda}^{\mathrm{T}}(C_0L_0C_0^{\mathrm{T}})^{-1}\hat{\lambda}. \tag{6.5.15}$$

定理 6.5.4 在原假设 H_0 和条件(C6.4.1)~条件(C6.4.5)下, 当 $n\to\infty$ 时, 则修正的 profile Lagrange 乘子检验统计量 $\hat{T}_{\mathrm{MPLMT}}$ 渐近收敛到标准 χ_k^2 分布, 其中 k 表示自由度, 即, $\hat{T}_{\mathrm{MPLMT}} \xrightarrow{L} \chi_k^2$.

定理 6.5.5 在对立假设 H_1 和条件(C6.4.1)~条件(C6.4.5)下, 当 $n\to\infty$ 时, 则修正的 profile Lagrange 乘子检验统计量 $\hat{T}_{\mathrm{MPLMT}}$ 渐近收敛到非中心的 $\chi^2(k,\varsigma)$ 分布, 其中 k 表示自由度, 非中心的参数为

$$\varsigma = \lim_{n\to\infty}(H\beta - d)^{\mathrm{T}}(CLC)^{-1}(H\beta - d),$$

其中 $C = \left(H\Sigma_1^{-1}H^{\mathrm{T}}\right)^{-1}$, $L = H\Sigma_1^{-1}\Sigma_2\Sigma_1^{-1}H^{\mathrm{T}}$.

注 6.5.4 定理 6.5.4 和定理 6.5.5 给出了修正的 profile Lagrange 乘子检验统计量的渐近分布. 定理不仅提供了半参数 EV 模型中关于参数分量的一种新的检验方法, 而且也证明了修正的 profile Lagrange 乘子检验统计量的 Wilks 现象. 由定理 6.5.4, 容易得到 $\hat{T}_{\mathrm{MPLMT}}$ 的渐近零分布与变量 T 的密度函数 $f(\cdot)$, 未知系数函数 $\alpha(\cdot)$, 以及参数 β 和 σ^2 无关. 因此, 可以通过渐近分布或者 Monte Carlo 模拟方法来计算渐近分布的临界值.

6.5.3 非参数部分协变量带有测量误差的情形

1. 估计方法

假设 $\{(Y_i; X_i^{\mathrm{T}}, W_i^{\mathrm{T}}, T_i), 1 \leqslant i \leqslant n\}$ 是来自模型 (6.4.16) 的独立同分布随机样本. 基于线性约束条件 (6.5.1), 可以定义 β 的调整 Lagrange 函数为

$$\begin{aligned} F(\beta,\lambda) = & \sum_{i=1}^{n}\{Y_i - X_i^{\mathrm{T}}\beta - W_i^{\mathrm{T}}\hat{\alpha}(T_i,\beta)\}^2 \\ & - \sum_{i=1}^{n}\hat{\alpha}^{\mathrm{T}}(T_i,\beta)\Sigma_{uu}\hat{\alpha}(T_i,\beta) + 2\lambda^{\mathrm{T}}(H\beta - d). \end{aligned} \tag{6.5.16}$$

令 $F(\beta,\lambda)$ 分别关于 β 和 λ 求导, 并令其导数等于 0, 则有

$$\left\{\begin{array}{l} \dfrac{\partial F(\beta,\lambda)}{\partial\beta} = -2\widetilde{X}^{\mathrm{T}}(\widetilde{Y} - \widetilde{X}\beta) + 2X^{\mathrm{T}}Q^{\mathrm{T}}I\otimes\Sigma_{uu}Q(Y - X\beta) + 2H^{\mathrm{T}}\lambda = 0, \\ \dfrac{\partial F(\beta,\lambda)}{\partial\lambda} = H\beta - d = 0. \end{array}\right. \tag{6.5.17}$$

对上式进行求解可得

$$\left\{\begin{array}{l} \hat{\beta}^{R_1} = \hat{\beta} - P^{-1}H^{\mathrm{T}}(HP^{-1}H^{\mathrm{T}})^{-1}(H\hat{\beta} - d), \\ \hat{\lambda} = (HP^{-1}H^{\mathrm{T}})^{-1}(H\hat{\beta} - d), \end{array}\right. \tag{6.5.18}$$

其中 $P = \widetilde{X}^{\mathrm{T}}\widetilde{X} - X^{\mathrm{T}}Q^{\mathrm{T}}I\otimes\Sigma_{uu}QX$, $\hat{\beta}$ 为式 (6.4.21) 定义的 β 的无约束 profile 局部纠偏最小二乘估计.

此外, 由 Lagrange 乘子法, 在线性约束条件 $H\beta = d$ 下, 关于 β 最小化 $(\hat{\beta} - \beta)^{\mathrm{T}}(\hat{\beta} - \beta)$ 可得 β 的另一种约束估计:

$$\hat{\beta}^{R_2} = \hat{\beta} - H^{\mathrm{T}}(HH^{\mathrm{T}})^{-1}(H\hat{\beta} - d). \tag{6.5.19}$$

称 $\hat{\beta}^{R_k}$, $k = 1, 2$ 为 β 的局部偏差纠正 profile 最小二乘估计 (BCRPLSE). 进一步, 由式 (6.4.20), 式 (6.5.18) 和式 (6.5.19) 可得系数函数 $\alpha(\cdot)$ 的估计为

$$\hat{\alpha}^{R_k}(T) = (I_q\ 0_q)\left\{(D_T^W)^{\mathrm{T}}\omega_T D_T^W - \Omega\right\}^{-1}(D_T^W)^{\mathrm{T}}\omega_T(Y - X\hat{\beta}^{R_k}), \quad k = 1, 2. \tag{6.5.20}$$

2. 渐近性质

定理 6.5.6 假设条件(C6.4.1)~条件(C6.4.4)和条件(C6.4.5′)成立, 当 $n \to \infty$ 时, 则有

$$\sqrt{n}(\hat{\beta}^{R_1} - \beta) \xrightarrow{L} N(0, \Xi),$$

其中

$$\Xi=\Xi_1^{-1}\Xi_2\Xi_1^{-1}-\Xi_1^{-1}\Xi_2\Xi_1^{-1}\Xi_3\Xi_1^{-1}-\Xi_1^{-1}\Xi_3\Xi_1^{-1}\Xi_2\Xi_1^{-1}+\Xi_1^{-1}\Xi_3\Xi_1^{-1}\Xi_2\Xi_1^{-1}\Xi_3\Xi_1^{-1},$$

$$\Xi_1=E(X_1X_1^{\mathrm T})-E(\Phi^{\mathrm T}(T_1)\Gamma^{-1}(T_1)\Phi(T_1)),\quad \Xi_3=H^{\mathrm T}[H\Xi_1^{-1}H^{\mathrm T}]^{-1}H,$$

$$\Xi_2=E(\varepsilon_1-u_1^{\mathrm T}\alpha(T_1))^2\Xi_1+\sigma^2E\left\{\Phi^{\mathrm T}(T_1)\Gamma^{-1}(T_1)\Sigma_{uu}\Gamma^{-1}(T_1)\Phi(T_1)\right\}^2$$
$$+E\left\{\Phi^{\mathrm T}(T_1)\Gamma^{-1}(T_1)\left(u_1u_1^{\mathrm T}-\Sigma_{uu}\right)\alpha(T_1)\right\}^{\otimes 2}.$$

定理 6.5.7 假设条件(C6.4.1)~条件(C6.4.4)和条件(C6.4.5′)成立, 当 $n\to\infty$ 时, 则有

$$\sqrt{n}(\hat\beta^{R_2}-\beta)\xrightarrow{L}N(0,\Pi),$$

其中 $\Xi_4=H^{\mathrm T}(HH^{\mathrm T})^{-1}H$,

$$\Pi=\Xi_1^{-1}\Xi_2\Xi_1^{-1}-\Xi_1^{-1}\Xi_2\Xi_1^{-1}\Xi_4-\Xi_4\Xi_1^{-1}\Xi_2\Xi_1^{-1}+\Xi_4\Xi_1^{-1}\Xi_2\Xi_1^{-1}\Xi_4.$$

为利用定理 6.5.6 和定理 6.5.7 对 β 进行统计推断, 需要估计渐近方差 Ξ 和 Π. 注意到, 只需估计 Ξ_1 和 Ξ_2, 进而由 plug-in 方法可得 Ξ 和 Π 的估计. Ξ_1 和 Ξ_2 的估计分别为

$$\hat\Xi_1=\frac{1}{n}\sum_{i=1}^n\widetilde X_i\widetilde X_i^{\mathrm T}-\frac{1}{n}\sum_{i=1}^nX^{\mathrm T}Q_i^{\mathrm T}\Sigma_{uu}Q_iX$$

和

$$\hat\Xi_2=\frac{1}{n}\sum_{i=1}^n\left\{\widetilde X_i(\widetilde Y_i-\widetilde X_i^{\mathrm T}\hat\beta)-\Sigma_{uu}\widetilde\alpha(T_i)\right\}^{\otimes 2}.$$

推论 6.5.2 假设条件(C6.4.1)~条件(C6.4.4)和条件(C6.4.5′)成立, 则在 Löwner 序意义下, $\hat\beta^{R_2}$ 比 $\hat\beta^{R_1}$ 更有效当且仅当 $\lambda_{\max}(\Pi\Xi^{-1})<1$, 其中 $\lambda_{\max}(\Pi\Xi^{-1})$ 是 $\Pi\Xi^{-1}$ 的最大特征值.

定理 6.5.8 假设条件(C6.4.1)~条件(C6.4.4)和条件(C6.4.5′)成立, 当 $n\to\infty$ 时, 则有

$$\sqrt{nh}\left(\hat\alpha^{R_k}(T)-\alpha(T)-\frac{h^2}{2}\frac{\mu_2^2-\mu_1\mu_3}{\mu_2-\mu_1^2}\alpha''(T)\right)\xrightarrow{L}N\Big(0,\Delta\Big),$$

其中

$$\Delta=(\kappa_0^2\nu_0+2\kappa_0\kappa_1\nu_1+\kappa_1^2\nu_2)f(T)^{-1}\Sigma^*,$$
$$\kappa_0=\mu_2/(\mu_2-\mu_1^2),\quad \kappa_1=-\mu_1/(\mu_2-\mu_1^2),$$
$$\Sigma^*=\Gamma^{-1}(T)[\sigma^2\Gamma(T)+\sigma^2\Sigma_{uu}+E\{\xi_1\alpha(T)\alpha^{\mathrm T}(T)\xi_1^{\mathrm T}|T_1=T\}]\Gamma^{-1}(T),$$
$$\xi_1=\Sigma_{uu}-u_1u_1^{\mathrm T}-X_1u_1^{\mathrm T}.$$

定理 6.5.9　假设条件(C6.4.1)～条件(C6.4.4)和条件(C6.4.5′)成立, 当 $n \to \infty$ 时, 则有

$$\max_{1\leqslant j\leqslant q}\sup_{T\in\mathcal{T}}|\hat{\alpha}_j^{R_k}(T)-\alpha_j(T)| = O\{h^2+(\log n/nh)^{\frac{1}{2}}\}\quad \text{a.s.}.$$

注 6.5.5　若 h 取最优窗宽, 即 $h_{\text{opt}} = cn^{-1/5}$, 其中 c 为一个常数, 则由定理 6.5.9 可知

$$\max_{1\leqslant j\leqslant q}\sup_{T\in\mathcal{T}}|\hat{\alpha}_j^{R_k}(T)-\alpha_j(T)| = O\left\{n^{-2/5}(\log n)^{1/2}\right\}\quad \text{a.s.}.$$

这说明模型 (6.4.16) 中非参数分量的估计的收敛速度达到了非参数函数估计的最优一致强收敛速度.

3. 假设检验

进一步, 考虑如下线性假设检验问题:

$$H_0: H\beta = d \quad \text{vs} \quad H_1: H\beta = d+\delta, \tag{6.5.21}$$

其中 δ 为 $k\times 1$ 常数向量. 对部分线性变系数模型, 当协变量 X 带有测量误差时, Zhang 等 (2011) 和 Wei (2012) 构造了上述检验问题的调整的 profile Lagrange 算子检验统计量. 这里考虑的是非参数部分协变量 Z 带有可加测量误差的情形. 为了消除测量误差对假设检验问题的影响, 提出了一个局部纠偏 profile Lagrange 算子检验方法, 从而使得所提检验统计量具有渐近标准卡方分布.

令 $G_0 = [H\hat{\Xi}_1^{-1}H^{\text{T}}]^{-1}$, $J_0 = H\hat{\Xi}_1^{-1}\hat{\Xi}_2\hat{\Xi}_1^{-1}H^{\text{T}}$, 其中 $\hat{\Xi}_1^{-1}\hat{\Xi}_2\hat{\Xi}_1^{-1}$ 是 $\Xi_1^{-1}\Xi_2\Xi_1^{-1}$ 的相合估计. 由式 (6.5.18), 可以构造如下局部偏差纠正 profile Lagrange 算子检验统计量

$$\hat{T}_{\text{BCPLM}} = \frac{1}{n}\hat{\lambda}^{\text{T}}(G_0J_0G_0^{\text{T}})^{-1}\hat{\lambda}.$$

定理 6.5.10　假设条件(C6.4.1)～条件(C6.4.4)和条件(C6.4.5′)成立, 当 $n \to \infty$ 时, 则有

(1) 当 H_0 成立时, 有 $\hat{T}_{\text{BCPLM}} \xrightarrow{L} \chi_k^2$;

(2) 当 H_1 成立时, $\hat{T}_{\text{BCPLM}}$ 渐近服从自由度为 k 的非中心卡方分布 $\chi^2(k,\varsigma)$, 且非中心参数为

$$\varsigma = \lim_{n\to\infty}(H\beta-d)^{\text{T}}G(GJG)^{-1}G(H\beta-d),$$

其中 $G = (H\Xi_1^{-1}H^{\text{T}})^{-1}$, $J = H\Xi_1^{-1}\Xi_2\Xi_1^{-1}H^{\text{T}}$.

6.5.4 模拟研究

例 6.5.1 考虑如下的部分线性变系数测量误差模型:

$$\begin{cases} Y_i = X_{i1}\beta_1 + X_{i2}\beta_2 + X_{i3}\beta_3 + Z_i\alpha(T_i) + \varepsilon_i, \\ V_{ij} = X_{ij} + e_{ij}, \quad j = 1, 2, 3, \end{cases}$$

其中 $X_{i1} \sim N(2,1)$, $X_{i2} \sim N(2,1)$, $X_{i3} \sim N(2,1)$, $Z_i \sim N(2,1)$, $T_i \sim U(0,1)$, $\beta = (1,1,1.5)^{\mathrm{T}}$, $\alpha(T_i) = \cos(2\pi T_i)$, 和 $\varepsilon_i \sim N(0, 0.5^2)$. 协变量 $X_i = (X_{i1}, X_{i2}, X_{i3})^{\mathrm{T}}$ 不能准确测量, 而具有测量误差, 且测量误差 $e_{ij} \sim N(0, \sigma_e^2)$, $j = 1, 2, 3$. 为了比较, 在模拟研究中考虑两种测量误差方差的情况: $\sigma_e^2 = 0.5$ 和 $\sigma_e^2 = 0.25$.

本例中将采用三个模拟来研究所提估计方法和检验程序的有限样本性质. 在模拟研究中, 使用高斯核函数 $K_h(\cdot) = \dfrac{1}{\sqrt{2\pi}h}\exp\big(-(\cdot)^2/2h^2\big)$. 窗宽 h 使用 CV 方法进行选取. 定义如下 h 的得分函数为

$$\mathrm{CV}(h) = \frac{1}{n}\sum_{i=1}^{n}\Big\{Y_i - V_i^{\mathrm{T}}\hat{\beta}_{-i} - Z_i^{\mathrm{T}}\hat{\alpha}_{h,-i}(T_i)\Big\}^2 - \hat{\beta}_{-i}^{\mathrm{T}}\Sigma_{ee}\hat{\beta}_{-i}, \tag{6.5.22}$$

其中 $\hat{\beta}_{-i}$ 是由去掉第 i 个个体后, 由式 (6.5.5) 所得到的修正的 profile 最小二乘估计, 且 $\hat{\alpha}_{h,-i}(\cdot)$ 是在式 (6.4.8) 中用 $\hat{\beta}_{-i}$ 代替 β 后所得到的估计. 交错核实的窗宽 h_{cv} 是最小化 $\mathrm{CV}(h)$ 的值, 即 $h_{\mathrm{cv}} = \arg\min\limits_{h}\mathrm{CV}(h)$.

在第一部分模拟中, 主要研究所提估计量 (MRPLSE) 的完成情况, 即比较估计量的偏差 (bias) 和均方误差 (MSE). 为了显示所提估计量 $\hat{\beta}^R$ 的有效性, 将它和另外的两种估计量进行比较: 忽略测量误差的估计量 (NE) $\check{\beta}^R$ 和修正的 profile 最小二乘估计 (MPLSE) $\hat{\beta}$. NE 估计量是直接忽略测量误差, 应用标准的 profile 最小二乘估计方法得到的约束估计量. 定义在式 (6.5.5) 中修正的 profile 最小二乘估计 (MPLSE) 是忽略约束条件而得到的估计量. 假设约束条件为 $H\beta = d$, 其中 $H = (0.5, 1, 1)$. 样本量大小分别取 $n = 100$, 200 和 300. 对每种情况, 重复模拟 1000 次, 并在两种不同测量误差下计算三种估计量 (MRPLSE, NE 和 MPLSE) 的偏差 (bias) 和均方误差 (MSE), 模拟结果报告在表 6.5.1 中. 此外, 当样本量大小为 200, 测量误差为 $\sigma_e^2 = 0.5$ 时, 图 6.5.1 给出了基于不同约束条件下非参数分量的拟合曲线.

表 6.5.1 说明有如下的模拟结果:

(1) 首先参数分量的 NE 估计是有偏的, 且偏差不会随着样本量的增大而变小, 而会随着测量误差方差的增大而变大;

表 6.5.1　参数分量估计的偏差和均方误差

方法		$\sigma_e^2=0.25$			$\sigma_e^2=0.5$		
		100	200	300	100	200	300
MRPLSE	$\hat{\beta}_1^R$	0.0074(0.0145)	0.0060(0.0063)	0.0036(0.0043)	0.0202(0.0349)	0.0128(0.0143)	0.0072(0.0096)
	$\hat{\beta}_2^R$	−0.0072(0.0108)	−0.0081(0.0050)	−0.0051(0.0031)	−0.0192(0.0276)	−0.0168(0.0117)	−0.0105(0.0073)
	$\hat{\beta}_3^R$	0.0035(0.0105)	0.0051(0.0047)	0.0032(0.0030)	0.0093(0.0255)	0.0104(0.0109)	0.0069(0.0069)
	AEE	0.0109	0.0113	0.0070	0.0294	0.0235	0.0145
NE	$\check{\beta}_1^R$	−0.0600(0.0125)	−0.0580(0.0074)	−0.0588(0.0062)	−0.1004(0.0237)	−0.0990(0.0160)	−0.1006(0.0143)
	$\check{\beta}_2^R$	0.0671(0.0106)	0.0634(0.0070)	0.0642(0.0060)	0.1110(0.0216)	0.1068(0.0158)	0.1078(0.0144)
	$\check{\beta}_3^R$	−0.0371(0.0074)	−0.0344(0.0040)	−0.0347(0.0030)	−0.0608(0.0127)	−0.0573(0.0075)	−0.0575(0.0060)
	AEE	0.0974	0.0926	0.0937	0.1615	0.1565	0.1583
MPLSE	$\hat{\beta}_1$	0.0037(0.0155)	0.0038(0.0066)	0.0023(0.0045)	0.0108(0.0392)	0.0077(0.0154)	0.0039(0.0103)
	$\hat{\beta}_2$	0.0088(0.0158)	0.0039(0.0070)	0.0016(0.0043)	0.0160(0.0406)	0.0065(0.0161)	0.0041(0.0101)
	$\hat{\beta}_3$	0.0205(0.0152)	0.0168(0.0068)	0.0101(0.0046)	0.0477(0.0403)	0.0350(0.0168)	0.0220(0.0108)
	AEE	0.0226	0.0177	0.0105	0.0515	0.0364	0.0227

注: MSE 见括号

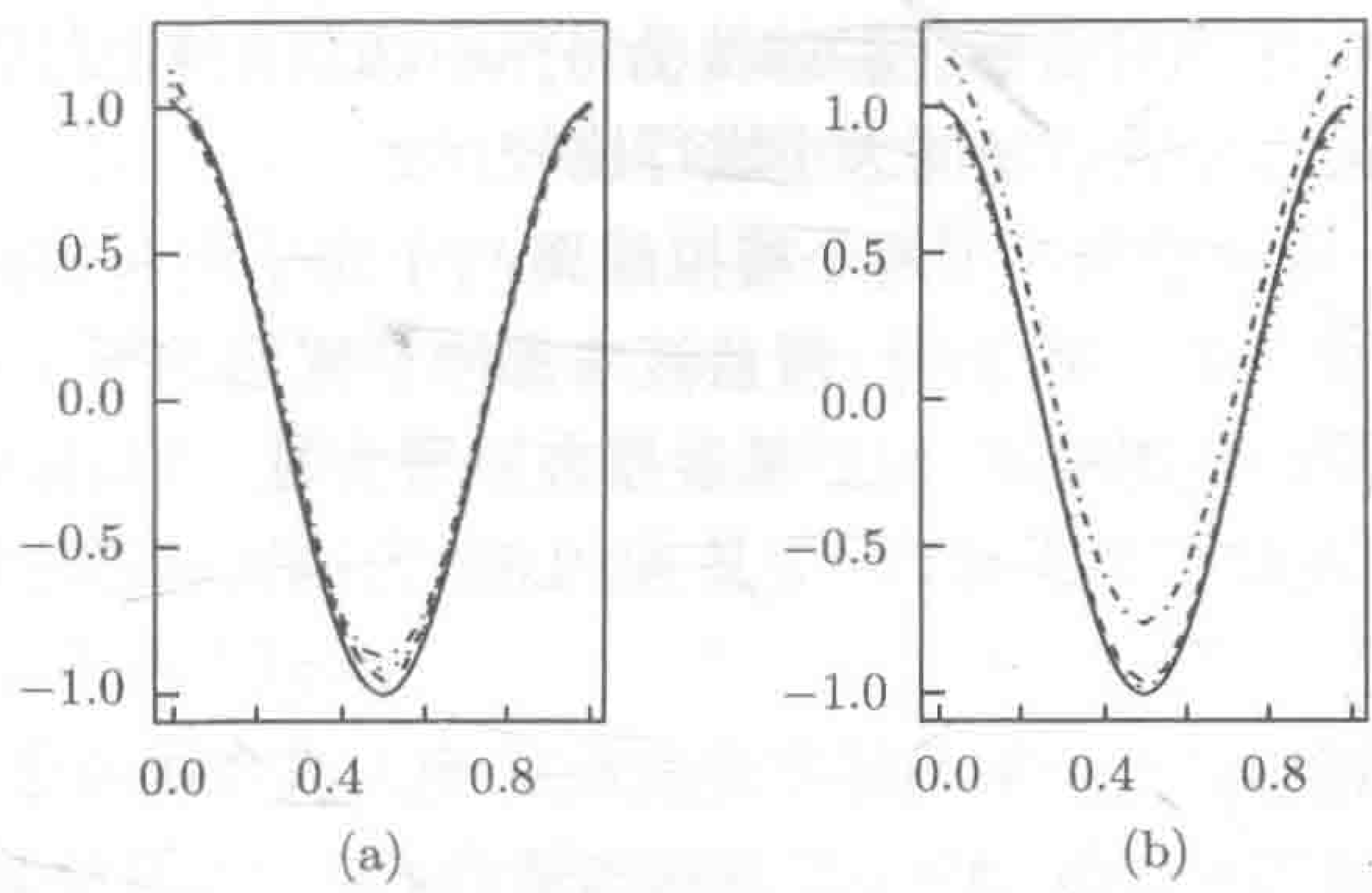


图 6.5.1 非参数分量在四种不同情形下的曲线拟合图, 其中点虚线表示 $\alpha(\cdot)$ 的 NE 估计, 虚线表示 $\alpha(\cdot)$ 的 MPLSE 估计, 点线表示 $\alpha(\cdot)$ 所提的 MRPLSE 估计, 实线表示真实曲线. (a) 图是基于约束条件 $H=(0.5,1,1)$ 的拟合曲线. (b) 图是基于约束条件 $H=(1,3,1)$ 的拟合曲线

(2) 对比 MRPLSE 和 MPLSE 两种估计的偏差, 能够发现参数分量的两种估计都是渐近无偏的. 由于表 6.5.1 中的偏差不能提供有效的比较信息, 进一步, 在表 6.5.1 中又计算了平均估计误差 (AEE), 即通过计算 L_2 范数 $\|\hat{\beta}-\beta\|$ 来比较两种方法估计的有效性. 发现 MPLSE 比 MRPLSE 有较大的平均估计误差 (AEE). 此外, 平均估计误差 (AEE) 随着样本量增大和测量误差变小而变小;

(3) 此外, 参数分量的 MRPLSE 比 MPLSE 有较小的均方误差 (MSE).

从图 6.5.1 可以看出所提出的非参数分量的估计 $\hat{\alpha}^R(\cdot)$ 的拟合曲线几乎和忽略约束条件的估计 $\hat{\alpha}(\cdot)$ 的拟合曲线重合, 并接近真实曲线, 而 NE 估计 $\check{\alpha}^R(\cdot)$ 是有偏的, 有较大的偏差. 这也说明所提出的非参数分量的估计是有效的.

在第二部分模拟中, 构造兴趣参数分量 β 的置信域. 为了构造置信域, 首先需要估计 $\hat{\beta}^R$ 的渐近方差. 此处采用 plug-in 方法来估计 $\hat{\beta}^R$ 的渐近方差. 为了简单起见, 仅计算了 β_3 的置信水平为 $1-\alpha=0.95$ 的置信区间和覆盖概率 (其他两个分量的结果也类似). 取样本量大小 $n=100,200$ 和 300, 比较平均覆盖概率 (CP) 和平均区间长度 (AL). 对于每种情况, 重复模拟次数为 1000 次, 模拟结果见表 6.5.2.

表 6.5.2 β_3 置信区间的平均长度 (AL) 和覆盖概率 (CP)

方法		$\sigma_e^2=0.25$			$\sigma_e^2=0.5$		
		100	200	300	100	200	300
MRPLSE	AL	0.3873	0.2687	0.2171	0.6104	0.4102	0.3277
	CP	0.9410	0.9370	0.9440	0.9450	0.9370	0.9450
MPLSE	AL	0.4651	0.3204	0.2577	0.7441	0.4927	0.3916
	CP	0.9340	0.9450	0.9460	0.9380	0.9450	0.9450

下面构造 (β_1,β_2) 的置信域. 在样本量大小为 $n=100$ 和 200 时, 分别计算其置

信水平为 $1-\alpha=0.95$ 的置信域, 图 6.5.2 为 $\sigma_e^2=0.25$ 的情况, 图 6.5.3 为 $\sigma_e^2=0.5$ 的情况, 模拟的结果可用置信域的大小进行比较.

(1) 从表 6.5.2 可以看出如下两个模拟结果: (i) 对于两种不同的测量误差和两种不同的方法, 当样本量 n 增加时, 覆盖概率都趋于置信水平 0.95, 且平均区间长度都变小. 对于相同的观测样本, 当测量误差方差增大时, 平均区间长度也变长; (ii) 尽管两种方法的覆盖概率非常接近, 但是 MPLSE 方法比所提方法 MRPLSE 有较长的平均区间长度.

(2) 图 6.5.2 和图 6.5.3 说明当样本量大小 n 变大或者测量误差方差变小时, 置信域将变小. 同时可以看出由 MPLSE 方法构造的 (β_1,β_2) 置信域比基于 MRPLSE 方法构造的置信域稍微大一些.

综上可知, 所提出的方法比忽略约束条件的修正的 profile 最小二乘方法有较好的估计效果.

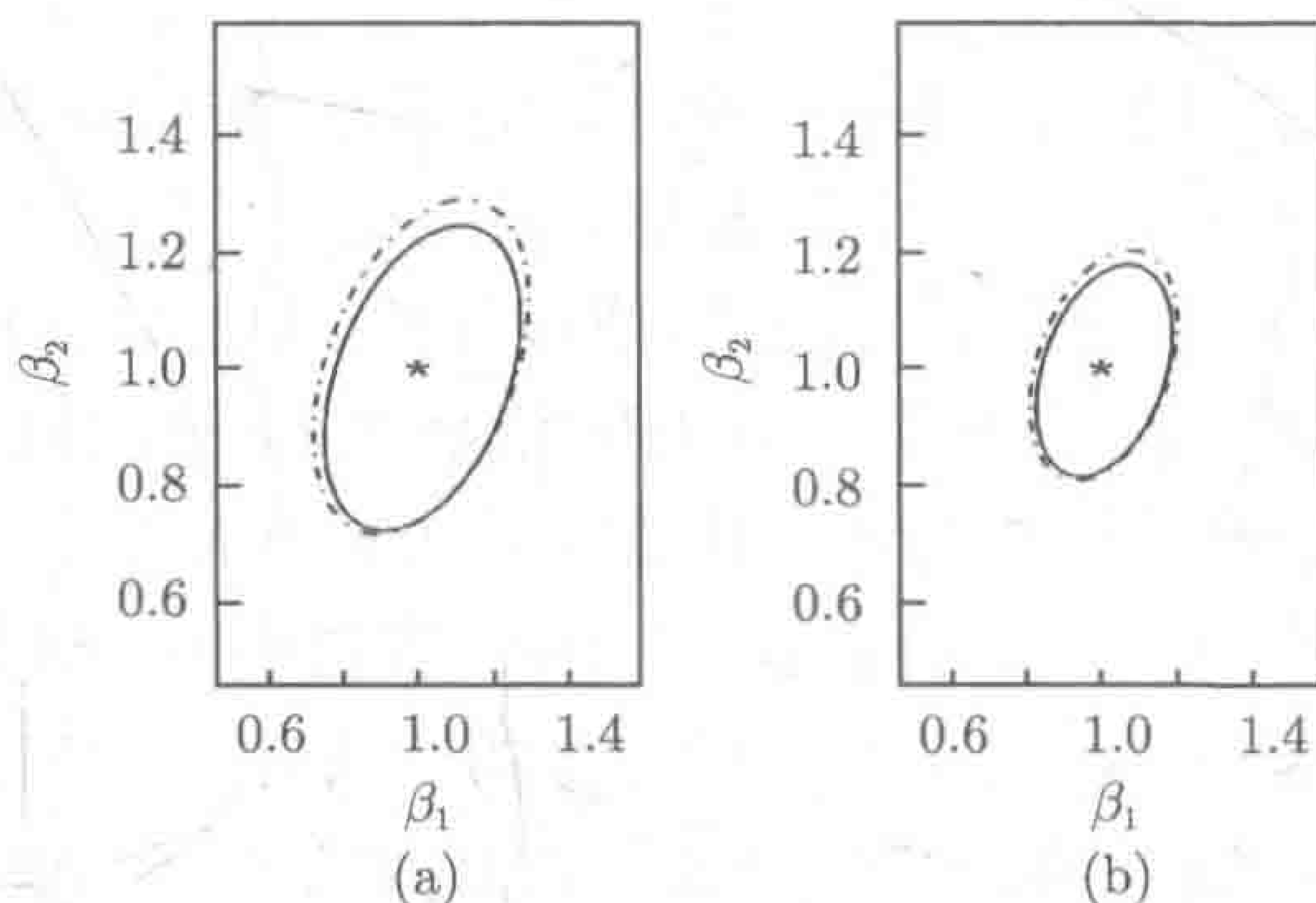


图 6.5.2　当 $\sigma_e^2=0.25$ 时, (β_1,β_2) 基于 MRPLSE 和 MPLSE 方法的置信域, 其中点虚线表示 MPLSE 方法, 实线表示 MRPLSE 方法. (a) 图表示样本量 $n=100$ 时的置信域, (b) 图表示 $n=200$ 时的置信域

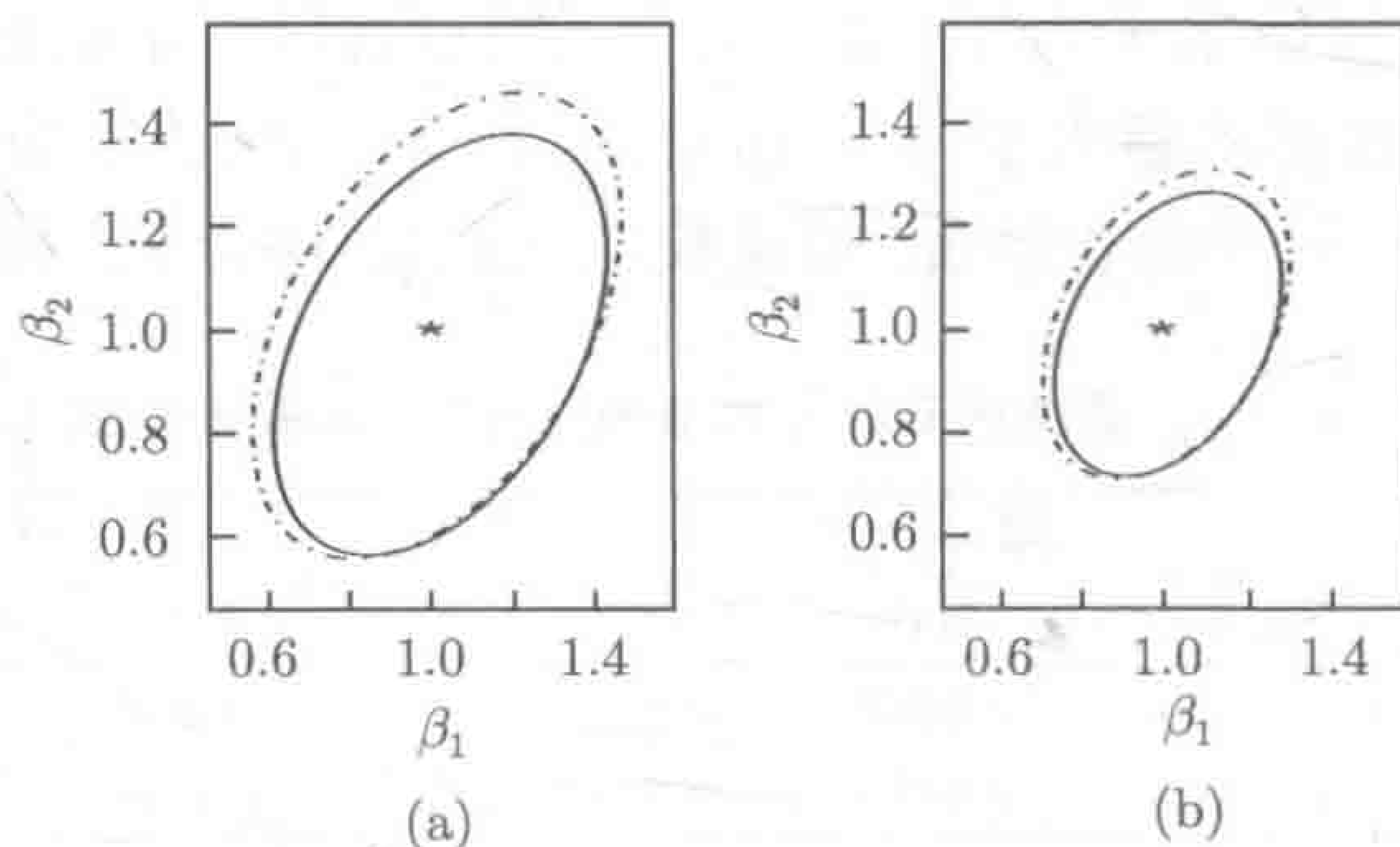


图 6.5.3　当 $\sigma_e^2=0.5$ 时, (β_1,β_2) 基于 MRPLSE 和 MPLSE 方法的置信域, 其中点虚线表示 MPLSE 方法, 实线表示 MRPLSE 方法. (a) 图表示样本量 $n=100$ 时的置信域, (b) 图表示 $n=200$ 时的置信域

在第三部分模拟中，主要考虑所提出的检验方法在有限样本下的表现. 取样本量大小分别为 $n = 100$ 和 $n = 200$. 考虑原假设为 $H\beta = d$, 其中 $H = (0.5, 1, 1)$, 相应的对立假设为 $H\beta = d + \delta$, 其中 δ 是一个正常数序列. 当 $\delta = 0$ 时, 对立假设即为原假设. 相应于定理 6.5.4, 修正的 profile Lagrange 乘子检验统计量是渐近到 χ^2 分布的. 基于该渐近分布, 可以确定临界值. 为了说明所提方法的有效性, 考虑两种测量误差情形: $\sigma_e^2 = 0.25$ 和 $\sigma_e^2 = 0.5$, 重复模拟 1000 次, 图 6.5.4 给出了显著性水平为 $\alpha = 0.05$ 的功效函数曲线.

从图 6.5.4 可以看出: ①当 δ 逐渐变大时, 估计的功效函数迅速增加, 这表明所提的检验统计量对备择假设 H_1 是相当敏感的; ② 对同一观测样本, 当测量误差的方差增大时, 估计的功效函数减小. 对于相同的测量误差, 当样本量增大时, 估计的功效函数增大. 可见测量误差的大小对功效函数是有影响的. 总体而言, 所提出的检验程序是可行的.

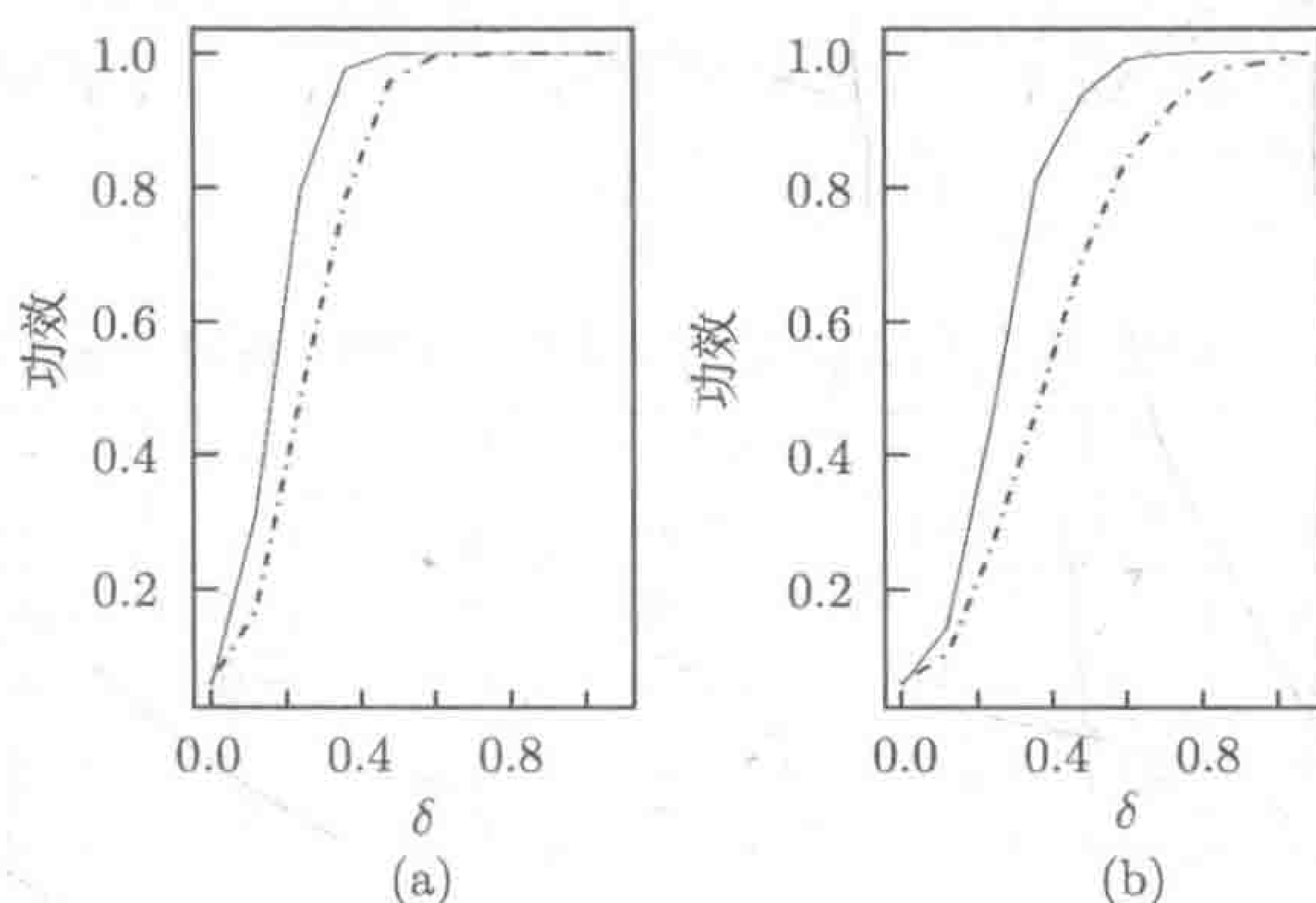


图 6.5.4 功效函数的曲线图, 其中点虚线表示 $n = 100$ 时的结果, 实线表示 $n = 200$ 时的结果. (a) 图是 $\sigma_e^2 = 0.25$ 时功效函数的曲线图, (b) 图是 $\sigma_e^2 = 0.5$ 时功效函数的曲线图

例 6.5.2 考虑如下的部分线性变系数测量误差模型:

$$\begin{cases} Y_i = X_{i1}\beta_1 + X_{i2}\beta_2 + X_{i3}\beta_3 + Z_i\alpha(T_i) + \varepsilon_i, \\ W_i = Z_i + u_i, \end{cases} \tag{6.5.23}$$

其中 $\beta = (1, 2, 1)^{\mathrm{T}}$, $\alpha(T_i) = 2\cos(2\pi T_i)$, $T_i \sim U(0, 1)$, $\varepsilon_i \sim N(0, 0.5^2)$, $(X_i^{\mathrm{T}}, Z_i)^{\mathrm{T}}$ 服从多元正态分布, 其均值为 $\mu = (1, 1, 1, 1)^{\mathrm{T}}$, 协方差阵 $\Sigma = (\sigma_{kl})$ 满足 $\sigma_{kl} = 0.5^{|k-l|}, k, l = 1, 2, 3, 4$. 进一步, 假定测量误差 $u_i \sim N(0, \sigma_u^2)$, 为更好地进行比较, 这里取两种测量误差方差: $\sigma_u^2 = 0.25$ 和 $\sigma_u^2 = 0.5$.

在下面的模拟中, 取 Epanechnikov 核函数 $K(t) = 0.75(1 - t^2)I(|t| \leqslant 1)$, 窗宽

h 采用 CV 方法选取. CV 统计量定义为:

$$\begin{aligned}\mathrm{CV}(h)=&\frac{1}{n}\sum_{i=1}^{n}\{Y_i-X_i^{\mathrm{T}}\hat{\beta}^{(-i)}-W_i^{\mathrm{T}}\hat{\alpha}_{h,(-i)}(T_i)\}^2\\&-\frac{1}{n}\sum_{i=1}^{n}\hat{\alpha}_{h,(-i)}^{\mathrm{T}}(T_i)\Sigma_{uu}\hat{\alpha}_{h,(-i)}(T_i),\end{aligned}\tag{6.5.24}$$

其中 $\hat{\beta}^{(-i)}$ 是去掉第 i 次观测后由式 (6.4.21) 得到的参数 β 的局部纠偏 profile 最小二乘估计, $\hat{\alpha}_{h,(-i)}(T_i)$ 是基于式 (6.4.20) 用 $\hat{\beta}^{(-i)}$ 代替 β 得到的估计. 最小化式 (6.5.24) 可得所需要的窗宽 h_{cv}, 即 $h_{\mathrm{cv}}=\arg\min\limits_{h>0}\mathrm{CV}(h)$.

首先, 取约束条件 $H\beta=d$, 其中 $H=(1,0.5,1)$, 比较两种约束估计 $\hat{\beta}^{R_1}$, $\hat{\beta}^{R_2}$ 和由式 (6.4.21) 得到的无约束估计 $\hat{\beta}$ 的均值 (Mean)、标准差 (SD) 和均方误差 (MSE). 样本容量 n 分别取 100, 200 和 400, 重复模拟计算 1000 次. 模拟结果见表 6.5.3. 另外, 当 $n=200$ 时, 在不同测量误差下比较了系数函数的三种估计: $\hat{\alpha}^{R_1}(t)$, $\hat{\alpha}^{R_2}(t)$ 和 $\hat{\alpha}^{NE}(t)$, 结果见图 6.5.5. 这里 $\hat{\alpha}^{NE}(t)$ 是忽略测量误差情形下基于 profile 最小二乘方法得到的系数函数的 naive 估计.

表 6.5.3 参数分量的约束及无约束估计的均值、标准差和均方误差

β	σ_u^2	n	$\hat{\beta}$			$\hat{\beta}^{R_1}$			$\hat{\beta}^{R_2}$		
			Mean	SD	MSE	Mean	SD	MSE	Mean	SD	MSE
β_1	0.25	100	1.0016	0.1264	0.0160	1.0095	0.1023	0.0105	1.0073	0.0981	0.0097
		200	0.9960	0.0739	0.0055	0.9956	0.0600	0.0036	0.9955	0.0606	0.0037
		400	0.9928	0.0511	0.0027	0.9961	0.0391	0.0015	0.9957	0.0390	0.0015
	0.5	100	1.0386	0.1180	0.0154	1.0154	0.1299	0.0171	1.0181	0.1273	0.0165
		200	0.9899	0.1138	0.0131	0.9806	0.0953	0.0095	0.9822	0.0981	0.0099
		400	1.0079	0.0508	0.0026	1.0024	0.422	0.0018	1.0031	0.0421	0.0018
β_2	0.25	100	2.0024	0.1338	0.0179	1.9944	0.1482	0.0220	2.0053	0.1475	0.0218
		200	2.0014	0.0820	0.0064	2.0020	0.0755	0.0057	2.0011	0.0894	0.0080
		400	2.0128	0.0550	0.0032	2.0096	0.0509	0.0027	2.0143	0.0626	0.0041
	0.5	100	2.0072	0.1673	0.0280	2.0242	0.1623	0.0269	1.9969	0.1798	0.0324
		200	2.0106	0.1294	0.0169	2.0177	0.1195	0.0146	2.0068	0.1445	0.0209
		400	1.9805	0.0536	0.0033	1.9828	0.0655	0.0046	1.9781	0.0539	0.0034
β_3	0.25	100	0.9844	0.1824	0.0335	0.9933	0.0872	0.0077	0.9901	0.1013	0.0104
		200	1.0045	0.0847	0.0072	1.0034	0.0541	0.0029	1.0040	0.0591	0.0035
		400	0.9943	0.0585	0.0035	0.9991	0.0359	0.0013	0.9972	0.0389	0.0015
	0.5	100	1.0039	0.2118	0.0449	0.9725	0.1258	0.0166	0.9834	0.1513	0.0232
		200	1.0220	0.1419	0.0206	1.0105	0.0807	0.0066	1.0144	0.0961	0.0095
		400	1.0126	0.0613	0.0039	1.0061	0.0307	0.0010	1.0079	0.0321	0.0011

表 6.5.4 参数分量的约束及无约束估计的平均估计误差

方法	$\sigma_u^2=0.25$			$\sigma_u^2=0.5$		
	$n=100$	$n=200$	$n=400$	$n=100$	$n=200$	$n=400$
$\|\hat{\beta}-\beta\|$	0.0673	0.0191	0.0092	0.0840	0.0500	0.0098
$\|\hat{\beta}^{R_1}-\beta\|$	0.0401	0.0122	0.0055	0.0577	0.0304	0.0066
$\|\hat{\beta}^{R_2}-\beta\|$	0.0418	0.0152	0.0071	0.0685	0.0399	0.0075

从表 6.5.3 的模拟结果可以看出: ① 测量误差方差 σ_u^2 的大小影响参数估计的精度, 当 σ_u^2 增大时, SD 和 MSE 均变大; ②随着样本量的增加, 所得三种参数估计的均值都越来越接近真实值, 且它们的 SD 和 MSE 均减小; ③对于所有情况, 参数分量的约束估计都优于相应的未约束估计. 另外, 由表 6.5.4, 通过计算平均估计误差 $\|\hat{\beta}^{R_1}-\beta\|$ 和 $\|\hat{\beta}^{R_2}-\beta\|$ 可以看出, 就此次模拟而言, $\hat{\beta}^{R_1}$ 比 $\hat{\beta}^{R_2}$ 更有效.

从图 6.5.5 可以看出, 系数函数的两种约束估计几乎重叠且都接近真实函数曲线, 说明所提局部偏差纠正方法是有效的. 此外, 系数函数的 naive 估计 $\hat{\alpha}^{NE}(t)$ 是有偏的, 且偏差随着测量误差方差的增加而变大.

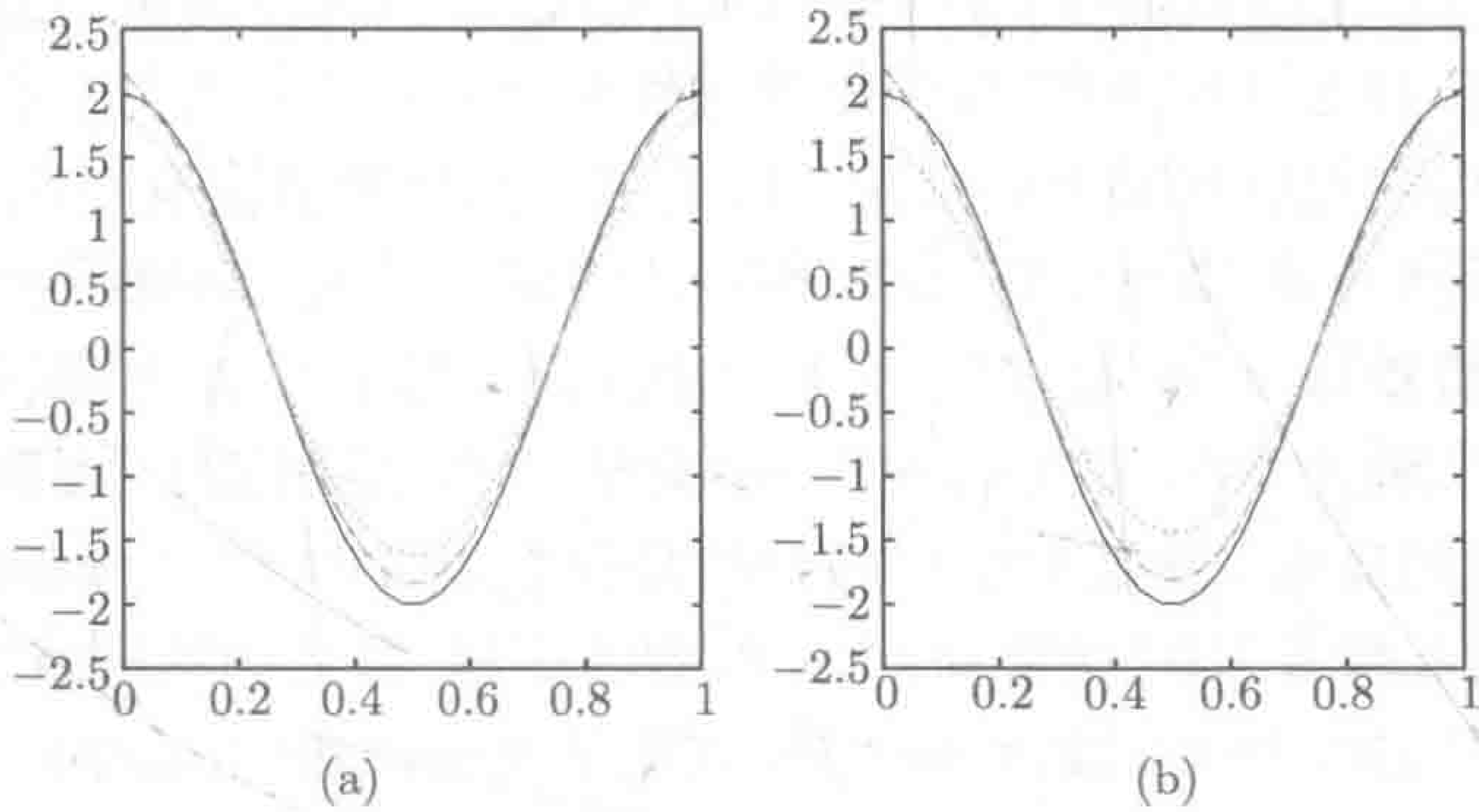


图 6.5.5 $n=200$ 时的模拟结果. 在每个图中, 实线为真实函数曲线, 点线为 $\hat{\alpha}^{NE}(t)$, 虚线为 $\hat{\alpha}^{R_1}(t)$, 点虚线为 $\hat{\alpha}^{R_2}(t)$, (a) 图是 $\sigma_u^2=0.25$ 的情形, (b) 图是 $\sigma_u^2=0.5$ 的情形

下面通过模拟来研究所提检验统计量的检验功效. 对模型 (6.5.23), 取原假设为 $H\beta=d$ 其中 $H=(1,0.5,1)$, 相应的备择假设为 $H\beta=d+\delta$ 其中 δ=0, 0.12, 0.24, $\cdots$, 1.08. 若 $\delta=0$, 则备择假设即转化为原假设. 取样本量 $n=200$, 显著性水平 $\alpha=0.05$, 重复计算 1000 次, 在 $\sigma_u^2=0.25$ 和 $\sigma_u^2=0.5$ 两种情形下, 图 6.5.6 给出了所提检验方法的检验功效函数.

从图 6.5.6 可以看出, 当原假设成立时 ($\delta=0$), 所提检验统计量的功效接近 0.05; 当备择假设成立时 ($\delta>0$), 检验功效随着 δ 的增加而快速变大. 这些结果说明所提出的检验统计量对备择假设是敏感的, 因此所提检验方法是可行的. 此外, 测量误差影响检验功效, 测量误差方差越小检验功效越大.

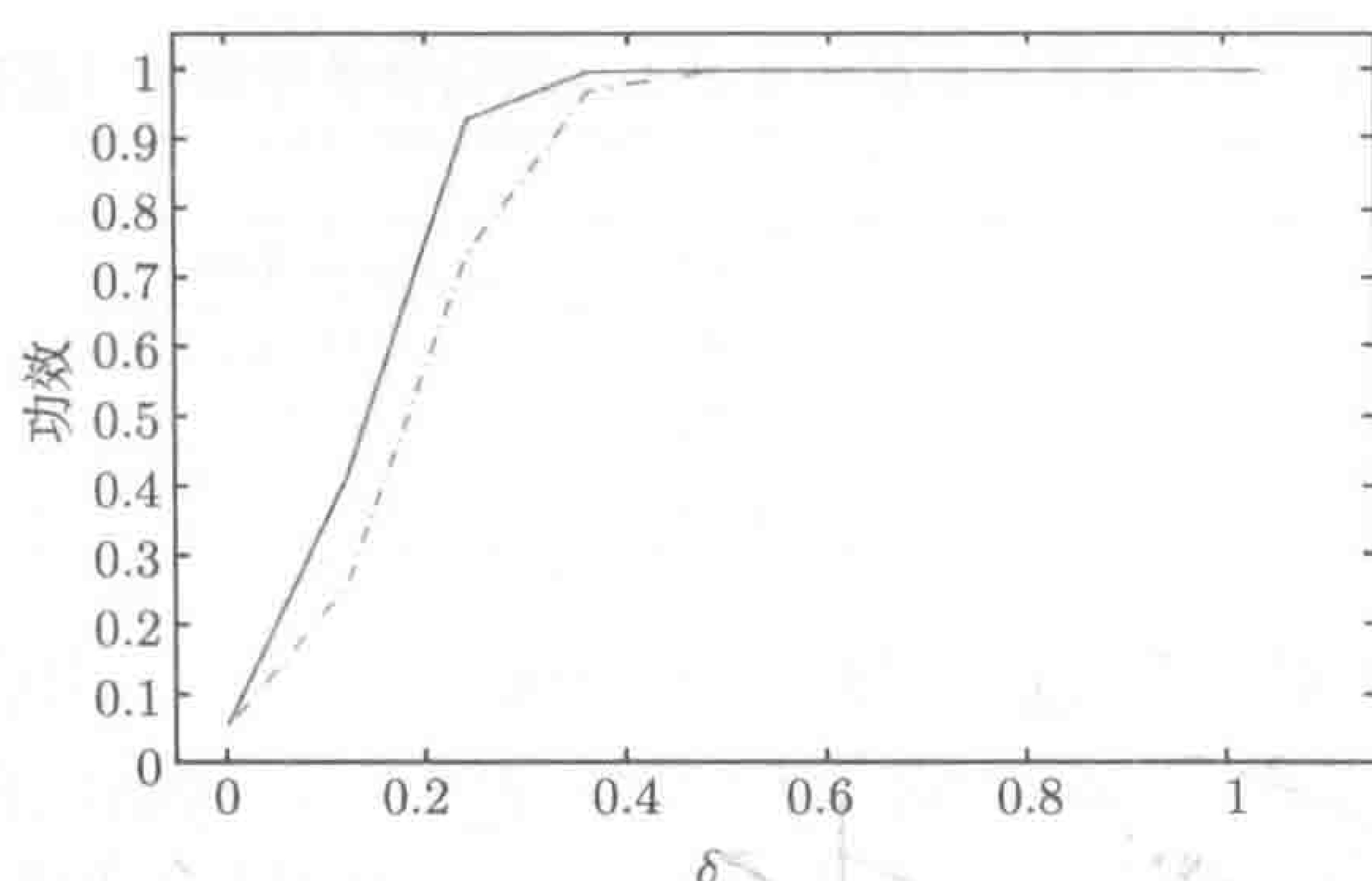


图 6.5.6　$n = 200$ 时的功效函数, 其中实线为基于 $\sigma_u^2 = 0.25$ 的模拟结果, 点虚线为基于 $\sigma_u^2 = 0.5$ 的模拟结果

6.5.5　实例分析

例 6.5.3(AIDS 数据分析)　为了说明 6.5.2 节所提估计方法的有效性, 本例对一组来自于多中心艾滋病群组 (AIDS) 研究的实际数据进行分析, 该数据是在 1984 年至 1991 年间, 对 283 位 HIV 呈阳性的同性恋患者每半年进行一次定期检查, 记录他们看病的医院地址和感染的情况. 但是由于部分患者没有定期来检查或者因病情发作而不到半年就需要检查一次, 每位患者重复测量的次数不同. 每位患者在这 8 年内至少检查过 1 次, 最多检查过 14 次. 对于这组数据, 响应变量是 HIV 感染后患者血液内所含 CD4 细胞的比例, 协变量是患者的年龄、吸烟状况、HIV 感染前 CD4 细胞的比例. 本数据研究的目的是探索患者的 CD4 细胞的比例随时间变化的趋势, 以及患者 HIV 感染前 CD4 细胞的比例, 患者的吸烟状况和 HIV 感染时患者的年龄对 HIV 感染后的患者 CD4 细胞的比例的影响. 许多统计学者使用变系数模型对该数据集进行了分析, 如 Huang 等 (2002), Xue 和 Zhu (2007), Wang 等 (2008). 他们考虑如下的变系数模型:

$$Y = \beta_0(T) + X_1\beta_1(T) + X_2\beta_2(T) + X_3\beta_3(T) + \varepsilon, \tag{6.5.25}$$

其中 Y 是患者的 CD4 细胞的比例, X_1 是患者 HIV 感染前 CD4 细胞的比例的中心化数据, X_2 是患者的吸烟状况, 取 1 表示吸烟, 取 0 表示非吸烟者, X_3 是患者感染 HIV 时中心化的年龄. 对模型 (6.5.25), $\beta_0(T)$ 是基准 CD4 细胞的比例函数, 表示感染 HIV 病毒 T 年后平均 CD4 细胞的比例. $\beta_1(T)$, $\beta_2(T)$ 和 $\beta_3(T)$ 分别表示患者感染 HIV 前 CD4 细胞比例对感染 HIV 后 CD4 细胞比例的影响效应, 吸烟对患者感染 HIV 后 CD4 细胞比例的效应和年龄的效应.

Huang 等 (2002) 的研究表明, 在显著性水平 0.05 下, 吸烟状况对患者的 CD4 细胞的比例没有显著性影响, 年龄对患者的 CD4 细胞的比例有较小的影响, 而患

者感染 HIV 前 CD4 细胞比例对感染 HIV 后 CD4 细胞比例的影响为常数效应, 不随时间变化. 根据 Huang 等 (2002) 的结论, 则模型 (6.5.25) 退化成部分线性变系数模型:

$$Y = \beta_0(T) + X_1\beta_1 + X_3\beta_3(T) + \varepsilon,$$

其中 β_1 是未知的参数. 对于该模型, 采用标准的 profile 最小二乘方法可得 β_1 的估计为 $\hat{\beta}_1 = 0.2096$. 事实上, 吸烟状况对患者的 CD4 细胞的比例有较小的影响, 则相应于 Huang 等 (2002) 的结论和先验信息, 因此考虑如下具有约束条件的部分线性变系数模型:

$$\begin{cases} Y = X_1\beta_1 + X_2\beta_2 + \beta_0(T) + X_3\beta_3(T) + \varepsilon, \\ H\beta = 0.2096, \end{cases} \tag{6.5.26}$$

其中 $H = (1, 15)$, $\beta = (\beta_1, \beta_2)^{\mathrm{T}}$ 是未知参数向量, $\beta_0(T)$ 和 $\beta_3(T)$ 是未知的系数函数. 首先把标准的约束 profile 最小二乘方法应用于模型 (6.5.26), 获得参数分量 β 的基准估计为 $\widetilde{\beta}^R = (0.1396, 0.0047)^{\mathrm{T}}$. 为了说明所提方法, 类似于 Lin 和 Carroll (2000), 此处作敏感性分析, 对变量 X_1 人为增加测量误差, 即

$$V_1 = X_1 + e_1, \tag{6.5.27}$$

其中 $e_1 \sim N(0, \sigma_e^2)$, 且考虑 $\sigma_e^2 = 0.25$ 和 $\sigma_e^2 = 0.5$ 两种情况进行比较. 这时把所提出的 MRPLSE 程序应用于模型 (6.5.26) 和 (6.5.27), 当测量误差的方差为 $\sigma_e^2 = 0.25$ 时, 获得参数分量 β 的估计为 $\hat{\beta}^R = (0.1498, 0.0039)^{\mathrm{T}}$, 当测量误差的方差为 $\sigma_e^2 = 0.5$ 时, 获得参数分量 β 的估计为 $\hat{\beta}^R = (0.1284, 0.0081)^{\mathrm{T}}$. 从该结果发现, 所得估计量非常接近于基准估计, 且患者感染 HIV 前 CD4 细胞比例有一个正的影响, β_2 的估计非常接近于 0. 此外, 基准函数 $\beta_0(T)$ 和年龄的系数函数 $\beta_3(T)$ 的估计曲线分别报告在图 6.5.7 和图 6.5.8 中.

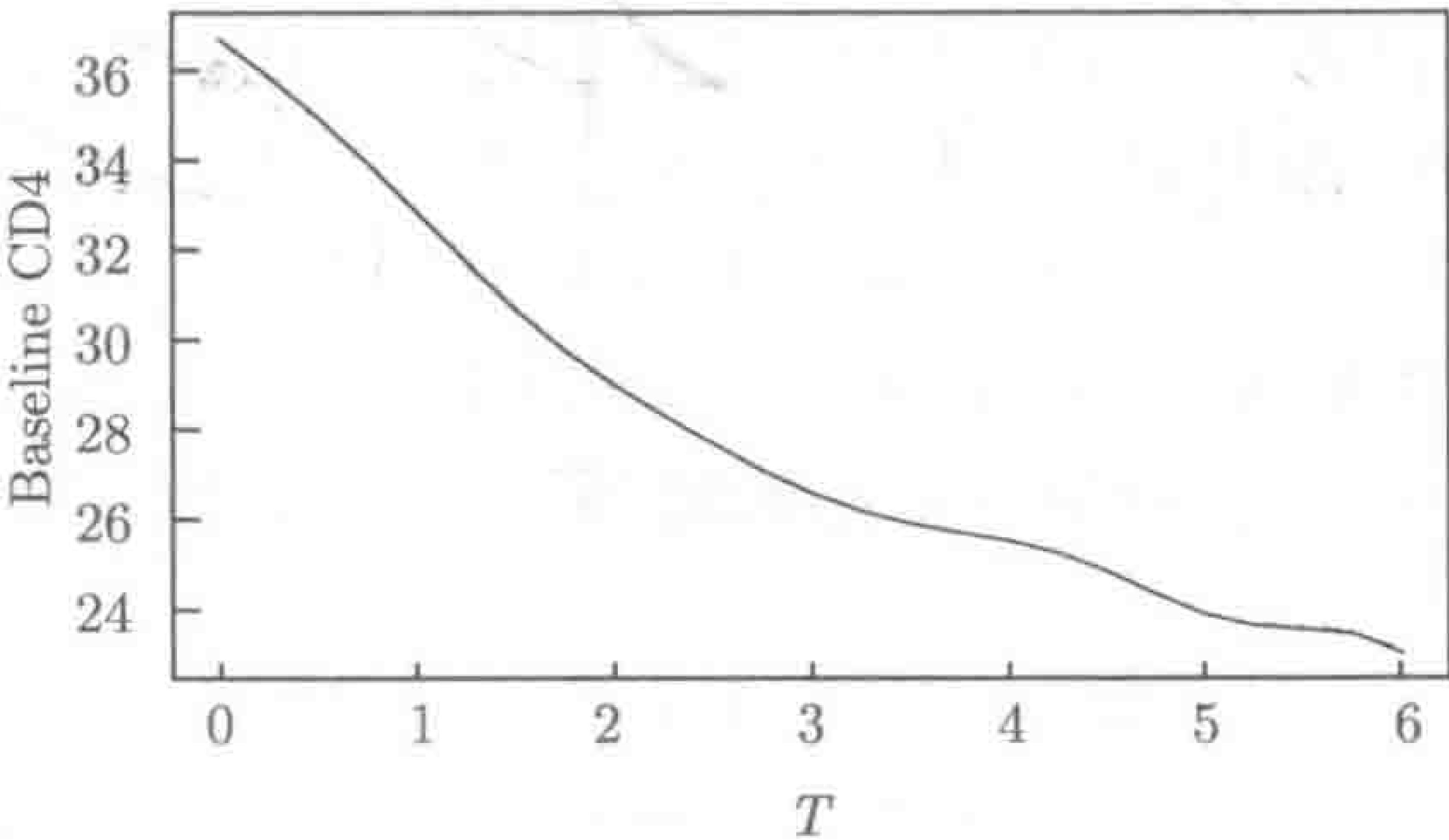


图 6.5.7 基于 BE 和 MRPLSE 两种方法得到的基准系数函数 $\beta_0(T)$ 的拟合曲线, 实线表示 BE 拟合曲线, 点虚线表示具有测量误差方差 $\sigma_u^2 = 0.25$ 的 MRPLSE 拟合曲线, 点线表示具有测量误差方差 $\sigma_u^2 = 0.5$ 的 MRPLSE 拟合曲线

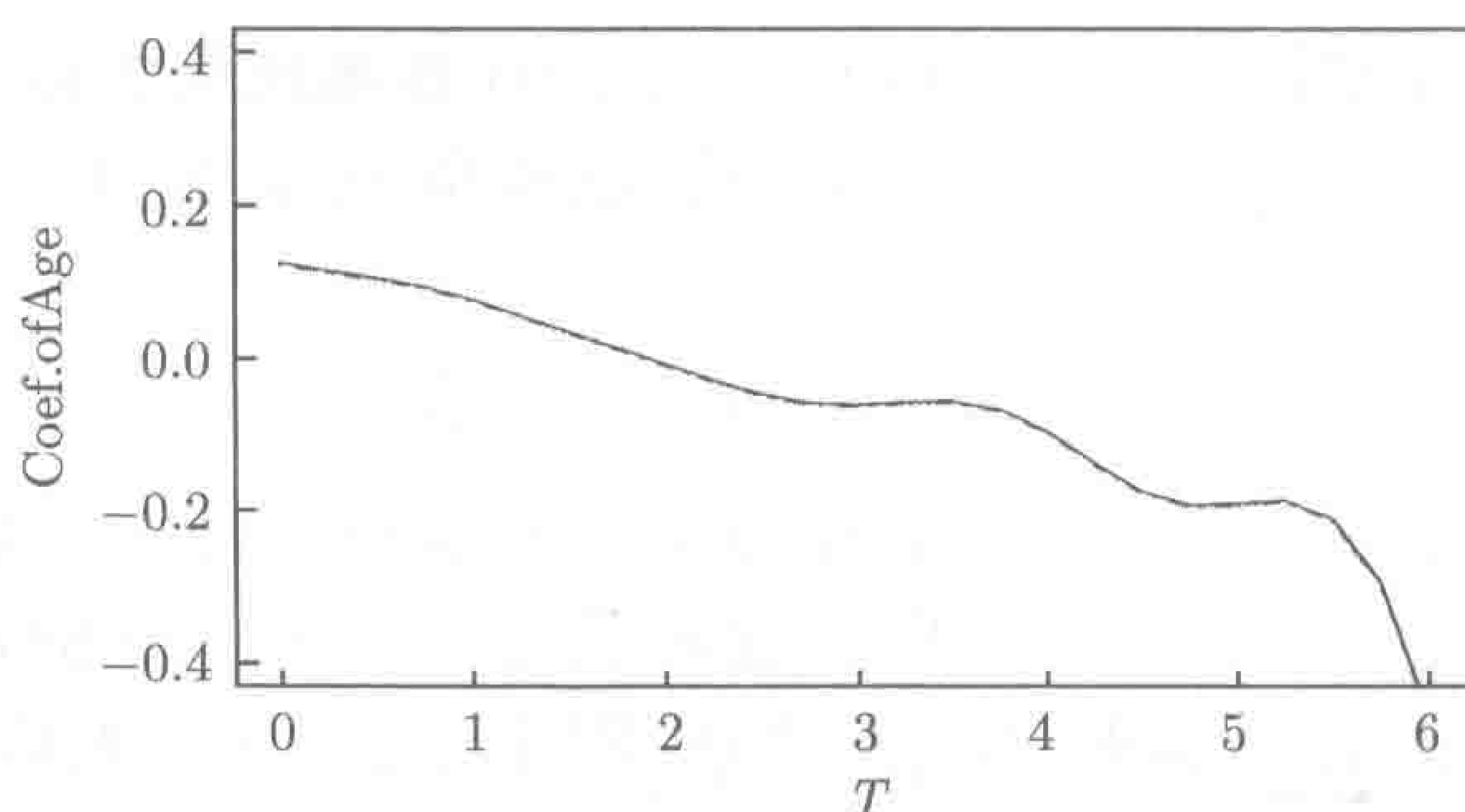


图 6.5.8 基于 BE 和 MRPLSE 两种方法得到的年龄系数函数 $\beta_3(T)$ 的拟合曲线, 实线表示 BE 拟合曲线, 点虚线表示具有测量误差方差 $\sigma_u^2 = 0.25$ 的 MRPLSE 拟合曲线, 点线表示具有测量误差方差 $\sigma_u^2 = 0.5$ 的 MRPLSE 拟合曲线

从图 6.5.7 中可以看出患者的平均 CD4 细胞比例在 HIV 感染初期下降非常迅速, 而在感染 4 年之后下降速度开始减缓. 从图 6.5.8 中可以看出年龄对患者感染后 CD4 细胞比例的影响效应趋于 0, 这也表明年龄对患者感染 HIV 后 CD4 细胞比例没有显著的影响. 这些发现与 Huang 等 (2002) 与 Xue 和 Zhu (2007) 利用变系数模型得到的结果是基本一致的. 图 6.5.7 和 6.5.8 也显示, 对不同测量误差情况, 系数函数 $\beta_0(T)$ 和 $\beta_3(T)$ 的 MRPLSE 程序的估计曲线和基准估计曲线是非常接近的, 从而表明所提出的纠偏方法是可行的, 所得理论结果是合理的.

例 6.5.4 (波士顿房产数据分析) 下面通过分析波士顿房产数据进一步说明 6.5.3 节所提局部纠偏约束估计方法的有效性. 波士顿房产数据源自 Harrison 和 Rubinfield (1978) 分析空气污染对住房价格影响的工作. 该数据集由住房价格和解释住房价格变化的一些变量构成, 其中住房价格为 1970 年居住在波士顿市区内的 506 个普查区域的自住型住房价格的中位数, 七个感兴趣的协变量分别为: 城市人均犯罪率 (CRIM)、每个住宅的房间平均数 (RM)、一氧化氮浓度 (NOX), 每 10000 美元含全部财产税 (TAX)、市区学校的师生比例 (PTRATIO)、建造于 1940 年之前的住房比例 (AGE) 以及底层居民的比例 (LSTAT). 为符号使用简单起见, 协变量 CRIM, RM, NOX, TAX, PTRATIO 和 AGE 分别记为 $Z_2, \cdots, Z_7$.

取 $Z_1 = 1$ 为截距项, 协变量 $T = \sqrt{\text{LSTAT}}$. Fan 和 Huang (2005) 讨论了 $Z_2, \cdots, Z_7$, 以及 LSTAT 对住房价格的影响, 并采用部分线性变系数模型:

$$Y = \sum_{i=1}^{5} \alpha_i(T) Z_i + \beta_1 Z_6 + \beta_2 Z_7 + \varepsilon \tag{6.5.28}$$

来拟合给定的数据. 为了说明 PTRATIO 和 AGE 的系数是否具有显著的统计意义, 他们分别采用广义似然比 (GLR) 和 Wald 统计量来进行检验, 结果显示 Z_7 在 1%的水平下是不显著的.

现在对模型 (6.5.28) 施加一个约束 $\beta_2 = 0$, 即取 $H = [0,1]$, $d = 0$. 在进行分析之前首先对响应变量和协变量 (Z_1 除外) 进行标准化变换, 同时也对变量 LSTAT 进行变换使其分布为 $U[0,1]$. 为证明所提方法的有效性, 类似于 Lin 和 Carroll (2000), 人工添加一个敏感性分析. 假定协变量 Z_5 带有测量误差, 即

$$W_5 = Z_5 + u_5, \tag{6.5.29}$$

其中 $u_5 \sim N(0, 0.3^2)$. 下面, 采用 6.5.3 节中所提的方法来估计模型 (6.5.28) 和模型 (6.5.29), 在约束条件 $H\beta = d$ 下, 可得参数 β 的 BCRPLSE 为 $\hat{\beta}^{R_1} = (-0.1479, 0)^{\mathrm{T}}$. 此外, β 的局部纠偏无约束估计为 $\hat{\beta} = (-0.1494, 0.0543)^{\mathrm{T}}$. 系数函数的估计曲线见图 6.5.9. 基于另外一种约束估计 $\hat{\beta}^{R_2}$ 的结果与上述结果类似, 这里略去.

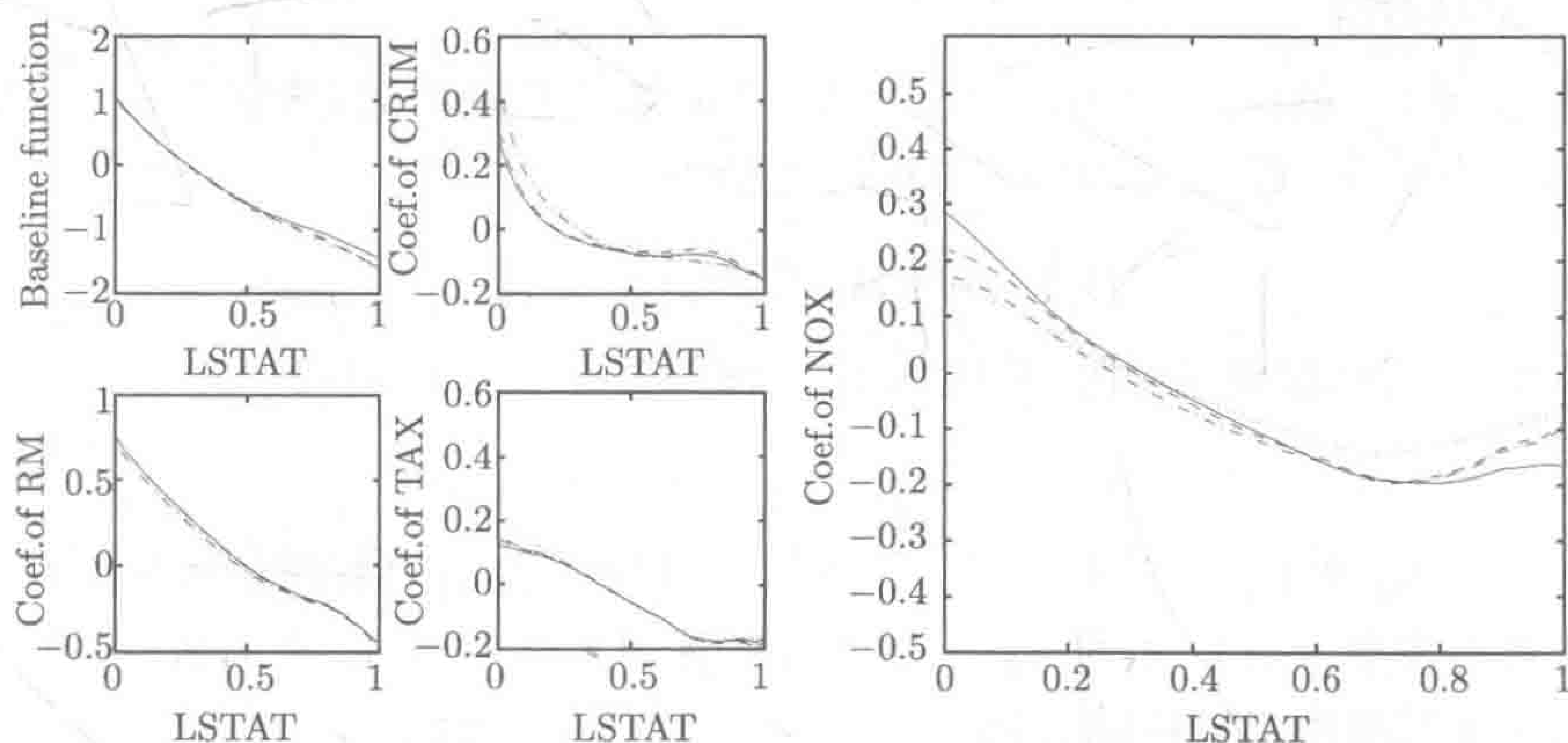


图 6.5.9 波士顿房产数据的应用. 实线为基准估计曲线, 点线为 naive 估计 $\hat{\alpha}^{NE}(\cdot)$, 虚线为 BCRPLSE 估计 $\hat{\alpha}^{R_1}(\cdot)$, 点虚线为无约束估计 $\tilde{\alpha}(\cdot)$

从图 6.5.9 可以看出: ① BCRPLSE 估计 $\hat{\alpha}^{R_1}(\cdot)$ 接近基于真实数据得到的基准估计; ② BCRPLSE 估计 $\hat{\alpha}^{R_1}(\cdot)$ 的表现优于 $\tilde{\alpha}(\cdot)$, 说明先验信息可以提高所提局部偏差纠正方法的估计效率; ③ 本节所提局部偏差纠正方法是有效的, BCRPLSE 估计 $\hat{\alpha}^{R_1}(\cdot)$ 和无约束估计 $\tilde{\alpha}(\cdot)$ 的表现都要优于忽略测量误差时得到的 naive 估计 $\hat{\alpha}^{NE}(\cdot)$.

6.6 部分线性变系数测量误差模型的变量选择

6.6.1 引言

考虑如下参数和非参数部分都带有测量误差的部分线性变系数模型:

$$\begin{cases} Y_i = X_i^{\mathrm{T}}\beta + Z_i^{\mathrm{T}}\alpha(T_i) + \varepsilon_i, \\ W_i = Z_i + u_i, \\ V_i = X_i + e_i, \quad i = 1, \cdots, n, \end{cases} \tag{6.6.1}$$

其中 Z_i 和 X_i 为不能精确观测的潜在协变量, W_i 和 V_i 为可以直接观测的替代变量. 模型误差 ε_i 独立同分布且有 $E(\varepsilon_i|X_i, Z_i, T_i) = 0$. u_i 和 e_i 为零均值测量误差, 且与 $(X_i, Z_i, T_i, \varepsilon_i)$ 独立. 考虑到模型的可识别性, 这里假定测量误差 u_i 和 e_i 的协方差阵 Σ_{uu} 和 Σ_{ee} 是已知的. 否则, 可以利用类似 2.2.1 节中提出的重复测量方法对其进行估计.

对于模型 (6.6.1), Zhao 和 Xue (2010) 利用基函数逼近以及惩罚最小二乘方法, 提出了一个偏差校正的变量选择方法. 进一步, 通过选择适当的截断参数, 证明了所提出的变量选择方法可以相合地识别出真实的模型, 且所得回归系数的正则估计具有 oracle 性质. 具体方法和结论是本节的主要内容.

6.6.2 估计方法

设 $B(u) = (B_1(u), \cdots, B_L(u))^{\mathrm{T}}$ 为 $M+1$ 阶 B 样条基函数, 其中 $L = K+M+1$, K 为内节点个数. 那么, $\alpha_k(t)$ 可以由下式逼近:

$$\alpha_k(t) \approx B^{\mathrm{T}}(t)\gamma_k, \quad k = 1, \cdots, q. \tag{6.6.2}$$

当 X_i 和 Z_i 可以精确观测时, 将式 (6.6.2) 代入模型 (6.6.1) 可得

$$Y_i \approx X_i^{\mathrm{T}}\beta + G_i^{\mathrm{T}}\gamma + \varepsilon_i, \quad i = 1, \cdots, n, \tag{6.6.3}$$

其中 $G_i = I_q \otimes B(T_i) \cdot Z_i$, $\gamma = (\gamma_1^{\mathrm{T}}, \cdots, \gamma_q^{\mathrm{T}})^{\mathrm{T}}$. 模型 (6.6.3) 为标准的线性回归模型, 并且注意到模型 (6.6.1) 中的 $\alpha_k(\cdot)$ 完全由模型 (6.6.3) 中的 γ_k 来刻画. 因此, 可以定义如下惩罚最小二乘目标函数:

$$Q(\beta, \gamma) = \sum_{i=1}^{n}\{Y_i - X_i^{\mathrm{T}}\beta - G_i^{\mathrm{T}}\gamma\}^2 + n\sum_{k=1}^{q} p_{\lambda_{1k}}(\|\gamma_k\|_H) + n\sum_{l=1}^{p} p_{\lambda_{2l}}(|\beta_l|), \tag{6.6.4}$$

其中 $\|\gamma_k\|_H = (\gamma_k^{\mathrm{T}} H \gamma_k)^{1/2}$, $H = \int B(t)B^{\mathrm{T}}(t)\mathrm{d}t$, $p_\lambda(\cdot)$ 为惩罚函数, λ 为截断参数. 此处仍采用 Fan 和 Li (2001) 提出的 SCAD 惩罚函数.

但是, 模型 (6.6.1) 中的 X_i 和 Z_i 均不能直接观测. 如果忽略测量误差并且直接用 V_i 和 W_i 分别代替 X_i 和 Z_i, 则可以证明所得的正则估计是不相合的. 另外, You 等 (2006) 针对变系数模型局部多项式估计的偏差较正方法将不能直接应用于这里的基于基函数逼近的估计方法. 于是, Zhao 和 Xue (2010) 提出一个新的偏差校正的方法. 记 $\Omega(t) = [I_q \otimes B(t)]\Sigma_{uu}[I_q \otimes B(t)]^{\mathrm{T}}$, 那么, 定义偏差校正的目标函数为

$$\begin{aligned} Q(\beta, \gamma) = & \sum_{i=1}^{n}\{Y_i - V_i^{\mathrm{T}}\beta - \widetilde{G}_i^{\mathrm{T}}\gamma\}^2 - \sum_{i=1}^{n}\gamma^{\mathrm{T}}\Omega(T_i)\gamma - n\beta^{\mathrm{T}}\Sigma_{ee}\beta \\ & + n\sum_{k=1}^{q} p_{\lambda_{1k}}(\|\gamma_k\|_H) + n\sum_{l=1}^{p} p_{\lambda_{2l}}(|\beta_l|), \end{aligned} \tag{6.6.5}$$

其中 $\widetilde{G}_i = I_q \otimes B(T_i) \cdot W_i$. 令 $\hat{\beta}$ 和 $\hat{\gamma} = (\hat{\gamma}_1^{\mathrm{T}}, \cdots, \hat{\gamma}_q^{\mathrm{T}})^{\mathrm{T}}$ 为最小化式 (6.6.5) 的解, 那么 $\hat{\beta}$ 为 β 的惩罚最小二乘估计, 且 $\alpha_k(t)$ 的估计可以定义为 $\hat{\alpha}_k(t) = B(t)^{\mathrm{T}}\gamma_k$.

6.6.3 算法

因为 $Q(\beta, \gamma)$ 中的惩罚函数在原点是奇异的, 所以普通的梯度算法将不能直接应用. 类似于 Fan 和 Li (2001), 利用对惩罚函数 $p_\lambda(\cdot)$ 进行局部二次逼近的方法给出一个迭代计算过程. 具体地, 在任一给定非零 ω_0 的某个邻域内, 惩罚函数 $p_\lambda(\cdot)$ 在 ω_0 点可以近似表示为

$$p_\lambda(|\omega|) \approx p_\lambda(|\omega_0|) + \frac{1}{2}\frac{p'_\lambda(|\omega_0|)}{|\omega_0|}(\omega^2 - \omega_0^2).$$

于是, 对给定的初始值 $\beta_l^{(0)}$ ($|\beta_l^{(0)}| > 0$), $l = 1, \cdots, p$, 和 $\gamma_k^{(0)}$ ($\|\gamma_k^{(0)}\|_H > 0$), $k = 1, \cdots, q$, 有

$$p_{\lambda_{2l}}(|\beta_l|) \approx p_{\lambda_{2l}}(|\beta_l^{(0)}|) + \frac{1}{2}\frac{p'_{\lambda_{2l}}(|\beta_l^{(0)}|)}{|\beta_l^{(0)}|}(|\beta_l|^2 - |\beta_l^{(0)}|^2), \tag{6.6.6}$$

$$p_{\lambda_{1k}}(\|\gamma_k\|_H) \approx p_{\lambda_{1k}}(\|\gamma_k^{(0)}\|_H) + \frac{1}{2}\frac{p'_{\lambda_{1k}}(\|\gamma_k^{(0)}\|_H)}{\|\gamma_k^{(0)}\|_H}(\|\gamma_k\|_H^2 - \|\gamma_k^{(0)}\|_H^2). \tag{6.6.7}$$

令

$$\Sigma(\beta^{(0)}) = \mathrm{diag}\left\{\frac{p'_{\lambda_{21}}(|\beta_1^{(0)}|)}{|\beta_1^{(0)}|}, \cdots, \frac{p'_{\lambda_{2p}}(|\beta_p^{(0)}|)}{|\beta_p^{(0)}|}\right\}$$

以及

$$\Sigma(\gamma^{(0)}) = \mathrm{diag}\left\{\frac{p'_{\lambda_{11}}(\|\gamma_1^{(0)}\|_H)}{\|\gamma_1^{(0)}\|_H}H, \cdots, \frac{p'_{\lambda_{1q}}(\|\gamma_q^{(0)}\|_H)}{\|\gamma_q^{(0)}\|_H}H\right\}.$$

那么, 结合式 (6.6.6) 和式 (6.6.7) 可知, 除了一个常数项, 式 (6.6.5) 可写为

$$\begin{aligned} Q(\beta, \gamma) = &\sum_{i=1}^n \{Y_i - V_i^{\mathrm{T}}\beta - \widetilde{G}_i^{\mathrm{T}}\gamma\}^2 - \sum_{i=1}^n \gamma^{\mathrm{T}}\Omega(T_i)\gamma - n\beta^{\mathrm{T}}\Sigma_{ee}\beta \\ &+ \frac{n}{2}\beta^{\mathrm{T}}\Sigma(\beta^{(0)})\beta + \frac{n}{2}\gamma^{\mathrm{T}}\Sigma(\gamma^{(0)})\gamma. \end{aligned} \tag{6.6.8}$$

注意到式 (6.6.8) 为一个二次型, 进而可以得到如下迭代计算过程.

步骤 1 初始化 $\beta^{(0)}$ 和 $\gamma^{(0)}$.

步骤 2 令 $\beta^{(0)} = \beta^{(k)}$ 和 $\gamma^{(0)} = \gamma^{(k)}$, 通过最小化式 (6.6.8) 求解 $\beta^{(k+1)}$ 和 $\gamma^{(k+1)}$.

步骤 3 重复计算步骤 2 直到收敛, 并记 β 和 γ 的最终估计为 $\hat{\beta}$ 和 $\hat{\gamma}$.

在初始化步, 可以利用式 (6.6.8) 右边的前三项, 并利用偏差校正的普通最小二乘方法给出 β 和 γ 的初始解. 为实施该算法, 内节点个数 K 和惩罚函数中的截断参数 λ_{1k}, λ_{2l} 和 a 需要选择. Fan 和 Li (2001) 表明模型的变量选择结果对 a 的选取不是太敏感, 并建议取 $a=3.7$. 取 $\lambda_{1k}=\lambda/\|\hat{\gamma}_k^{(0)}\|_H$, $\lambda_{2l}=\lambda/|\hat{\beta}_l^{(0)}|$, 其中 $\hat{\gamma}_k^{(0)}$ 和 $\hat{\beta}_l^{(0)}$ 分别为 γ_k 和 β_l 的初始估计. 另外, 通过最小化如下交错核实得分函数来选择 λ 和 K.

$$\mathrm{CV}(K,\lambda)=\sum_{i=1}^{n}\{Y_i-V_i^{\mathrm{T}}\hat{\beta}_{[i]}-\widetilde{G}_i^{\mathrm{T}}\hat{\gamma}_{[i]}\}^2-\sum_{i=1}^{n}\hat{\gamma}_{[i]}^{\mathrm{T}}\Omega(T_i)\hat{\gamma}_{[i]}-\sum_{i=1}^{n}\hat{\beta}_{[i]}^{\mathrm{T}}\Sigma_{ee}\hat{\beta}_{[i]}, \quad (6.6.9)$$

其中 $\hat{\beta}_{[i]}$ 和 $\hat{\gamma}_{[i]}$ 为去掉第 i 个个体后, 式 (6.6.5) 的解.

Zhao 和 Xue (2010) 在一定正则条件下证明了所得参数和非参数分量的惩罚最小二乘估计的相合性、稀疏性和渐近正态性. 关于定理的详细讨论和证明, 读者可参考 Zhao 和 Xue (2010).

第 7 章 单指标测量误差模型

本章重点考虑单指标模型和部分线性单指标模型, 在可加测量误差和受失真函数污染后的测量误差两种情况下, 研究模型中参数和非参数分量的估计方法和理论结果.

7.1 引 言

7.1.1 部分线性单指标模型及研究现状

1.1.5 节已经简单介绍过单指标模型, 该模型是 20 世纪 80 年代中后期发展起来的一种重要的统计模型, 已在金融、经济、生物和医学等各个领域广泛应用. 本章考虑下面的部分线性单指标模型:

$$Y = X^{\mathrm{T}}\beta + g(Z^{\mathrm{T}}\theta) + \varepsilon, \tag{7.1.1}$$

其中 β 是 $p \times 1$ 维的未知参数向量, θ 是 $q \times 1$ 维的未知指标参数向量, X 和 Z 分别是 p 维和 q 维的协变量, Y 是响应变量, ε 是方差有限的模型误差, 且满足 $E(\varepsilon|X,Z) = 0$, $g(\cdot)$ 是未知的一元光滑联系函数. 此外, 为了使得指标参数向量 θ 与未知联系函数 $g(\cdot)$ 能够被识别, 不失一般性, 通常假设 θ 的模长为 1 并且它的第一个分量规定为正数, 即 θ 的参数空间可规定为 $\theta = \{\theta = (\theta_1, \theta_2, \cdots, \theta_q)^{\mathrm{T}}, \|\theta\| = 1, \theta_1 > 0, \theta \in \mathbb{R}^q\}$.

部分线性单指标模型 (7.1.1) 包括两类重要的半参数回归模型: 当协变量 Z 的维数为 1 时, 模型 (7.1.1) 就退化为第 5 章中提到的部分线性模型; 当参数向量 $\beta = 0$ 时, 模型 (7.1.1) 就变为单指标模型 (Härdle et al., 1993; Horowitz, 2009; Horowitz and Härdle, 1996; Ichimura, 1993). 关于单指标模型的研究已经取得了可喜的研究成果, 薛留根 (2012a) 的专著《现代统计模型》第 3 章对单指标模型的估计方法和理论结果进行了详细的概括和总结, 薛留根 (2012b, 2012c) 对单指标模型研究的发展进行了综述, 李高荣和杨宜平 (2015) 的专著《纵向数据半参数模型》第 5 章和第 6 章对纵向数据下的单指标模型和部分线性单指标模型的估计方法和理论结果进行了详细的介绍, 这些都是对单指标模型了解和学习的最好资料, 感兴趣的读者可参考.

下面介绍关于单指标模型和部分线性单指标模型近几年的一些最新研究成果.

Guo 等 (2015) 考虑了在协变量缺失情况下, 关于指标参数构造了逆概率加权经验似然的统计推断方法. Wang 和 Wang (2015) 考虑了模型误差在平稳和强混合条件下, 研究了指标参数在发散维数情形下的估计和变量选择问题. Zhu 等 (2015) 关于异方差单指标模型提出了两种核光滑的非参数模型检验方法, 并研究了检验统计量的渐近性质. Zhang 等 (2015a) 结合 Xia 等 (2002) 提出的最小平均方差估计 (minimum average variance estimation, MAVE) 的思想, 给出了条件误差分布函数中指标参数的估计方法并研究了相关估计量的渐近性质. Lai 等 (2013a) 针对纵向数据部分线性单指标模型, 为了避免估计组内相关结构和提高估计的效率, 提出了纠偏的二次推断函数方法, 证明了即使错误指定相关结构, 所得到的估计量仍然是渐近有效的. 此外, 他们进一步把广义似然比检验方法推广到纵向数据部分线性单指标模型, 检验模型中的联系函数是否是线性结构. Li 等 (2015) 进一步把广义估计方程方法和二次推断函数方法应用到纵向数据部分线性单指标模型, 提出了惩罚的纠偏广义估计方程和惩罚的纠偏二次推断函数的变量选择方法, 并证明了两种变量选择方法的相合性和 oracle 性质. Ma 等 (2014) 基于二次推断函数的样条方法将部分线性单指标模型用于分析重复测量数据, 研究了指标参数和联系函数的渐近性质, 同时研究了变量选择问题. Lai 等 (2013b) 研究了面板数据单指标固定效应模型的半参数估计问题, 证明了估计量的渐近正态性. 对于单指标模型中联系函数同时置信带的研究, Li 等 (2014a) 在独立同分布数据下构造了单指标模型中联系函数的置信带, 并研究了联系函数的拟合优度检验问题. Yang 等 (2014) 把 Li 等 (2014a) 的同时置信带的构造方法推广到纵向数据单指标混合效应模型的研究, 构造了联系函数的同时置信带, 并建立了相应的理论结果. Ding 等 (2015) 基于 MAVE 算法提出了转换部分线性单指标模型中指标参数和联系函数的估计方法. Lian 等 (2015) 构造了部分线性单指标条件均值模型和部分线性单指标条件方差模型, 提出了模型中参数的有效估计方法, 并研究了所得估计量的渐近性质. Wang 和 Shi (2014) 基于局部线性核光滑 M 估计方法, 将部分线性单指标模型用于分析带有异常值的生存数据, 给出了指标参数的估计方法并给出了联系函数估计的最优收敛速度. Poon 和 Wang (2014) 研究了部分线性单指标模型中响应变量 Y 是多维情形下估计方法, 基于贝叶斯样条估计方法给出了指标参数和联系函数的估计方法. Chen 和 Parker (2014) 讨论了部分线性单指标模型中指标参数估计的半参数有效性问题. Cui 等 (2012) 研究了广义单指标模型, 提出了不动点迭代算法, 讨论了广义单指标模型中指标参数估计的有效性问题.

下面介绍部分线性单指标模型的两种估计方法, 分别是 Xia 和 Härdle (2006) 进一步发展了 Xia 等 (2002) 提出的 MAVE 方法和 Zhang 等 (2012c) 提出的降维方法.

7.1.2 MAVE 方法

Xia 和 Härdle (2006) 把 Xia 等 (2002) 提出的 MAVE 方法进行推广, 下面首先介绍 MAVE 的主要思想. 模型 (7.1.1) 中的参数 (β,θ) 的真实参数, 可通过极小化如下约束的目标函数得到, 定义为

$$(\beta_0,\theta_0)=\arg\min_{\beta,\theta}E\left[Y-\{X^{\mathrm{T}}\beta+g(Z^{\mathrm{T}}\theta)\}\right]^2,\quad \|\theta\|=1. \tag{7.1.2}$$

给定条件 $\xi=Z^{\mathrm{T}}\theta$ 下, 式 (7.1.2) 定义的解等价于极小化 $E_\xi\sigma^2_{\beta,\theta}(\xi)$, 其中

$$\sigma^2_{\beta,\theta}(\xi)=E\left[\left(Y-\{X^{\mathrm{T}}\beta+g(\xi)\}\right)^2\middle|Z^{\mathrm{T}}\theta=\xi\right] \tag{7.1.3}$$

为条件方差. 容易推得

$$E\left[Y-\{X^{\mathrm{T}}\beta+g(Z^{\mathrm{T}}\theta)\}\right]^2=E_\xi\sigma^2_{\beta,\theta}(\xi). \tag{7.1.4}$$

因此, 式 (7.1.2) 定义的解等价于

$$(\beta_0,\theta_0)=\arg\min_{\beta,\theta}E_\xi\sigma^2_{\beta,\theta}(\xi),\quad \|\theta\|=1. \tag{7.1.5}$$

因此, Xia 等 (2002) 与 Xia 和 Härdle (2006) 把这种极小化平均条件方差所得到的估计称为最小平均方差估计 (MAVE).

假设 $\{(Y_i,X_i,Z_i),i=1,\cdots,n\}$ 是 i.i.d. 的随机样本, 来自于模型 (7.1.1). 式 (7.1.5) 定义的条件期望可以由样本进行逼近. 对于 Z_i 在 z 的一个邻域内, 有下面的局部线性逼近:

$$\begin{aligned}Y_i-X_i^{\mathrm{T}}\beta-g(Z_i^{\mathrm{T}}\theta)&\approx Y_i-X_i^{\mathrm{T}}\beta-g(z^{\mathrm{T}}\theta)-g'(z^{\mathrm{T}}\theta)\theta^{\mathrm{T}}(Z_i-z)\\&=:Y_i-X_i^{\mathrm{T}}\beta-a-d\theta^{\mathrm{T}}(Z_i-z).\end{aligned} \tag{7.1.6}$$

由局部线性光滑的思想, 能够估计 $\sigma^2_{\beta,\theta}(z^{\mathrm{T}}\theta)$ 如下:

$$\hat\sigma^2_{\beta,\theta}(z^{\mathrm{T}}\theta)=\min_{a,d}\sum_{i=1}^n\left\{Y_i-X_i^{\mathrm{T}}\beta-a-d\theta^{\mathrm{T}}(Z_i-z)\right\}^2\omega_{i0}, \tag{7.1.7}$$

其中 $\omega_{i0}\geqslant 0,i=1,\cdots,n$ 是一些权函数, 满足 $\sum\limits_{i=1}^n\omega_{i0}=1$. 基于上面的讨论, 下面介绍详细的 MAVE 算法.

第一步, 先用高维的核密度权函数的核估计方法得到指标参数 θ 的相合估计. 若数据 Z_i 和 Z_j 距离比较接近时, 将 $g(Z_i^{\mathrm{T}}\theta)$ 在 $Z_j^{\mathrm{T}}\theta$ 的邻域内进行如下的局部线性逼近

$$\begin{aligned}g(Z_i^{\mathrm{T}}\theta)&\approx g(Z_j^{\mathrm{T}}\theta)+g'(Z_j^{\mathrm{T}}\theta)\theta^{\mathrm{T}}(Z_i-Z_j)\\&=:a_j+d_j\theta^{\mathrm{T}}(Z_i-Z_j).\end{aligned}$$

定义核密度权函数 $\omega_i(Z_j)=\dfrac{I_n(Z_j)H_{h_1}(Z_i-Z_j)}{\sum\limits_{i=1}^{n}H_{h_1}(Z_i-Z_j)}$, 其中 $H(\cdot)$ 是一个 q 维变量的多元密度函数, $H_{h_1}(\cdot)=h_1^{-q}H(\cdot/h_1)$, h_1 是窗宽. 示性函数定义为

$$I_n(Z_j)=I\left\{\frac{1}{n}\sum_{i=1}^{n}H_{h_1}(Z_i-Z_j)>c_0\right\},$$

其中 c_0 是一个预先给定的常数, 示性函数的作用是为了消除估计量的边界效应.

将未知量 $d_j\theta$ 看成一个整体, 关于 $\{\beta,a_j,d_j\theta\}$ 极小化下面的加权最小二乘目标函数

$$\sum_{i=1}^{n}\sum_{j=1}^{n}\left\{Y_i-X_i^{\mathrm{T}}\beta-a_j-(d_j\theta)^{\mathrm{T}}(Z_i-Z_j)\right\}^2\omega_i(Z_j). \tag{7.1.8}$$

不难看出, 极小化目标函数 (7.1.8) 可通过分别对 $j=1,\cdots,n$ 极小化

$$\sum_{i=1}^{n}\left\{Y_i-X_i^{\mathrm{T}}\beta-a_j-(d_j\theta)^{\mathrm{T}}(Z_i-Z_j)\right\}^2\omega_i(Z_j)$$

来完成. 具体如下:

$$\begin{aligned}\begin{pmatrix}\hat{\beta}_{(j)}\\ \hat{a}_j\\ \widehat{d_j\theta}\end{pmatrix}=&\left[\sum_{i=1}^{n}\omega_i(Z_j)\begin{pmatrix}X_i\\ 1\\ Z_i-Z_j\end{pmatrix}\begin{pmatrix}X_i\\ 1\\ Z_i-Z_j\end{pmatrix}^{\mathrm{T}}\right]^{-1}\\ &\times\left[\sum_{i=1}^{n}\omega_i(Z_j)\begin{pmatrix}X_i\\ 1\\ Z_i-Z_j\end{pmatrix}Y_i\right].\end{aligned} \tag{7.1.9}$$

记 $\hat{\mathcal{V}}_n=\left(\widehat{d_1\theta},\widehat{d_2\theta},\cdots,\widehat{d_n\theta}\right)^{\mathrm{T}}$, 其中 $\widehat{d_j\theta}$ 是式 (7.1.9) 所得到的估计, $j=1,\cdots,n$. 定义 $\hat{\varsigma}$ 是矩阵 $\hat{\mathcal{V}}_n\hat{\mathcal{V}}_n^{\mathrm{T}}/n$ 最大特征值所对应的特征向量. 由于真值 θ 的第一个分量为正数, 那么可定义θ 的初始估计为 $\hat{\theta}_{\mathrm{ini}}=\mathrm{sgn}(\hat{\varsigma}_1)\hat{\varsigma}$, 其中 $\hat{\varsigma}_1$ 是 $\hat{\varsigma}$ 的第一个分量, $\mathrm{sgn}(\cdot)$ 是符号函数. 参数向量 β 的初始估计可定义为 $\hat{\beta}_{\mathrm{ini}}=\dfrac{1}{n}\sum\limits_{j=1}^{n}\hat{\beta}_{(j)}$, 其中$\hat{\beta}_{(j)}$ 是由式 (7.1.9) 所得到的估计, $j=1,\cdots,n$. 因此, 第一步主要目的是获得 (β,θ) 的初始估计, 记为 $(\hat{\beta}_{\mathrm{ini}},\hat{\theta}_{\mathrm{ini}})$.

窗宽是一个很重要的光滑参数, 会直接影响参数的初始估计效果, 因为窗宽选择一直是非参数估计的一个重要问题. 这里采用去一分量交错核实方法来选取窗宽

h_1. 定义

$$\mathrm{CV}_0(h_1)=\sum_{j=1}^{n}\left\{Y_j-X_j^{\mathrm{T}}\hat{\beta}_{(j)}^{\backslash j}-\hat{a}_j^{\backslash j}\right\}^2,$$

其中 $\hat{\beta}_{(j)}^{\backslash j}$ 和 $\hat{a}_j^{\backslash j}$ 由下式所定义:

$$\begin{pmatrix}\hat{\beta}_{(j)}^{\backslash j}\\ \hat{a}_j^{\backslash j}\\ \widehat{d_j\theta}^{\backslash j}\end{pmatrix}=\left[\sum_{i=1,i\neq j}^{n}\omega_i(Z_j)\begin{pmatrix}X_i\\ 1\\ Z_i-Z_j\end{pmatrix}\begin{pmatrix}X_i\\ 1\\ Z_i-Z_j\end{pmatrix}^{\mathrm{T}}\right]^{-1}$$
$$\times\left[\sum_{i=1,i\neq j}^{n}\omega_i(Z_j)\begin{pmatrix}X_i\\ 1\\ Z_i-Z_j\end{pmatrix}Y_i\right].\tag{7.1.10}$$

极小化 $\mathrm{CV}_0(h_1)$, 则得到第一步所需的最优窗宽 h_1, 定义为 $\hat{h}_1=\arg\min\limits_{h_1}\mathrm{CV}_0(h_1)$.

第二步, Xia 和 Härdle (2006) 构造了未知参数 (β,θ) 和未知联系函数 $g(\cdot)$ 之间相互迭代的 MAVE 算法. 构造如下的目标函数

$$\sum_{i=1}^{n}\sum_{j=1}^{n}\left\{Y_i-X_i^{\mathrm{T}}\beta-a_j-(d_j\theta)^{\mathrm{T}}(Z_i-Z_j)\right\}^2\omega_i(Z_j^{\mathrm{T}}\theta),\tag{7.1.11}$$

其中$\omega_i(Z_j^{\mathrm{T}}\theta)=\dfrac{I_n(Z_j^{\mathrm{T}}\theta)K_h(Z_i^{\mathrm{T}}\theta-Z_j^{\mathrm{T}}\theta)}{\sum\limits_{i=1}^{n}K_h(Z_i^{\mathrm{T}}\theta-Z_j^{\mathrm{T}}\theta)}$, $K_h(\cdot)=K(\cdot/h)/h$, $K(\cdot)$ 是 1 维随机变量的密度函数, h 是窗宽. 示性函数

$$I_n(Z_j^{\mathrm{T}}\theta)=I\left\{\frac{1}{n}\sum_{i=1}^{n}K_h(Z_i^{\mathrm{T}}\theta-Z_j^{\mathrm{T}}\theta)>c_1\right\},$$

其中 c_1 是一个预先给定的常数. 具体算法如下.

步骤 1　给定 (β,θ), 关于 a_j 和 d_j 极小化目标函数 (7.1.11), 可得到

$$\begin{pmatrix}\hat{a}_j\\ \hat{d}_j\end{pmatrix}=\left[\sum_{i=1}^{n}\omega_i(Z_j^{\mathrm{T}}\theta)\begin{pmatrix}1\\ \theta^{\mathrm{T}}(Z_i-Z_j)\end{pmatrix}\begin{pmatrix}1\\ \theta^{\mathrm{T}}(Z_i-Z_j)\end{pmatrix}^{\mathrm{T}}\right]^{-1}$$
$$\times\left[\sum_{i=1}^{n}\omega_i(Z_j^{\mathrm{T}}\theta)\begin{pmatrix}1\\ \theta^{\mathrm{T}}(Z_i-Z_j)\end{pmatrix}(Y_i-X_i^{\mathrm{T}}\beta)\right].\tag{7.1.12}$$

步骤 2　给定 (a_j,d_j), 关于 β 和 θ 极小化目标函数 (7.1.11), 可得到

$$\begin{aligned}\begin{pmatrix}\hat{\beta}\\ \hat{\theta}\end{pmatrix}=&\left[\sum_{i=1}^{n}\sum_{j=1}^{n}\omega_i(Z_j^{\mathrm{T}}\theta)\begin{pmatrix}X_i\\ d_j(Z_i-Z_j)\end{pmatrix}\begin{pmatrix}X_i\\ d_j(Z_i-Z_j)\end{pmatrix}^{\mathrm{T}}\right]^{-1}\\ &\times\left[\sum_{i=1}^{n}\sum_{j=1}^{n}\omega_i(Z_j^{\mathrm{T}}\theta)\begin{pmatrix}X_i\\ d_j(Z_i-Z_j)\end{pmatrix}(Y_i-a_j)\right].\end{aligned}\tag{7.1.13}$$

进一步, 正则化 $\hat{\theta}=\mathrm{sgn}(\hat{\theta}_1)\hat{\theta}/\|\hat{\theta}\|$, 其中 $\hat{\theta}_1$ 是 $\hat{\theta}$ 的第一个分量.

在第二步中, 先用第一步得到的初始估计 $(\hat{\beta}_{\mathrm{ini}},\hat{\theta}_{\mathrm{ini}})$, 代入到式 (7.1.12) 中得到 a_j 和 d_j 的估计, 分别记为 $\hat{a}_j$ 和 $\hat{d}_j$. 然后把 $\hat{a}_j$ 和 $\hat{d}_j$ 代入步骤 2 中的式 (7.1.13) 中, 得到 β 和 θ 的更新估计. 不断地在步骤 1 和步骤 2 之间进行迭代, 在某种收敛准则下停止迭代, 并将相对应的最终收敛值作为参数 β 和 θ 的最终估计. Xia 和 Härdle (2006) 讨论了参数 (β,θ) 估计量的相合性和渐近正态性.

在第二步中, 窗宽 h 的选取同样采用去一分量交错核实方法. 定义

$$\mathrm{CV}(h)=\sum_{j=1}^{n}\left\{Y_j-X_j^{\mathrm{T}}\hat{\beta}-\hat{a}_j^{\backslash j}\right\}^2,$$

其中 $\hat{\beta}$ 可由式 (7.1.13) 的迭代值所确定, 而 $\hat{a}_j^{\backslash j}$ 由下式所定义:

$$\begin{aligned}\begin{pmatrix}\hat{a}_j^{\backslash j}\\ \hat{d}_j^{\backslash j}\end{pmatrix}=&\left[\sum_{i=1,i\neq j}^{n}\omega_i(Z_j^{\mathrm{T}}\hat{\theta})\begin{pmatrix}1\\ \hat{\theta}^{\mathrm{T}}(Z_i-Z_j)\end{pmatrix}\begin{pmatrix}1\\ \hat{\theta}^{\mathrm{T}}(Z_i-Z_j)\end{pmatrix}^{\mathrm{T}}\right]^{-1}\\ &\times\left[\sum_{i=1,i\neq j}^{n}\omega_i(Z_j^{\mathrm{T}}\hat{\theta})\begin{pmatrix}1\\ \hat{\theta}^{\mathrm{T}}(Z_i-Z_j)\end{pmatrix}\left(Y_i-X_i^{\mathrm{T}}\hat{\beta}\right)\right],\end{aligned}\tag{7.1.14}$$

其中 $(\hat{\beta},\hat{\theta})$ 是由式 (7.1.13) 所定义的迭代收敛估计. 极小化 $\mathrm{CV}(h)$, 获得最优的窗宽 h 为 $\hat{h}=\arg\min\limits_{h}\mathrm{CV}(h)$.

7.1.3　降维方法

Xia 和 Härdle (2006) 提出的 MAVE 方法在协变量 X 和 Z 的维数较高的情况下, 估计效果不是很理想. 因为在第一步 (β,θ) 的初始估计中, 需要高维核密度函数, 类似于非参数估计, 这样会有较慢的收敛速度, 直接影响 (β,θ) 的初始估计效率. 为了解决该问题, Zhang 等 (2012c) 提出了一种基于降维的计算方法, 该方法适用于协变量维数较高的情形. Zhang 等 (2012c) 在假设 (X,Z) 和模型误差 ε 相互独立的情况下, 将模型 (7.1.1) 转换为如下的降维结构

$$Y \perp\!\!\!\perp (X^{\mathrm{T}},Z^{\mathrm{T}})^{\mathrm{T}}\,\Big|\,\mathcal{S}^{\mathrm{T}}(X^{\mathrm{T}},Z^{\mathrm{T}})^{\mathrm{T}},\tag{7.1.15}$$

其中

$$\mathcal{S} = \begin{pmatrix} \beta & 0_{p\times 1} \\ 0_{q\times 1} & \theta \end{pmatrix},$$

这里 $\perp\!\!\!\perp$ 表示独立. 假设 $\{(Y_i, X_i, Z_i), i=1,\cdots,n\}$ 是来自于模型 (7.1.1) 的 i.i.d. 的随机样本, 具体估计方法如下.

步骤 1 基于选定的降维方法给出 $c_1\theta$ 和 $c_2\beta$ 的估计, 其中 c_1 和 c_2 是不为零的常数. 将该估计量分别记为 $\hat{v}_1$ 和 $\hat{v}_2$, 即 $\widehat{c_1\theta}=\hat{v}_1$ 和 $\widehat{c_2\beta}=\hat{v}_2$.

步骤 2 正则化 $\hat{v}_1$ 得到 θ 的估计: $\hat{\theta}=\dfrac{\hat{v}_1}{\|\hat{v}_1\|}\mathrm{sgn}(\hat{v}_{1,1})$, 其中 $\hat{v}_{1,1}$ 是 $\hat{v}_1$ 的第一个分量.

步骤 3 将步骤 2 得到的估计 $\hat{\theta}$ 代入到模型 (7.1.1) 中, 转换为如下的部分线性模型:

$$Y_i \approx \kappa\left(\hat{v}_2^{\mathrm{T}} X_i\right) + g\left(\hat{\theta}^{\mathrm{T}} Z_i\right) + \varepsilon_i. \tag{7.1.16}$$

这时, 可利用部分线性模型的估计方法 (见 5.1.1 节), 给出参数 κ 与联系函数 $g(\cdot)$ 的估计, 分别记为 $\hat{\kappa}$ 和 $\hat{g}(\cdot)$. 最后, β 的估计为 $\hat{\beta}=\hat{\kappa}\hat{v}_2$.

在步骤 1 中使用的降维方法, 可选用 Zhu 等 (2010) 提出的累积分布降维方法或 Li (1991) 提出的切片逆回归方法. 简单介绍 Zhu 等 (2010) 提出的累积分布降维方法. Zhang 等 (2012c) 给出了一个识别性条件, 并证明在该识别性条件下使用累积分布降维方法得到的最大特征值所对应的特征向量 η 可以识别 $c_1\theta$ 和 $c_2\beta$. 即 η 的前 p 个分量与 β 成比例关系, η 的后 q 个分量与 θ 成比例关系. 在总体意义下, 可进行如下的特征值分解:

$$\Sigma^{-1/2} M \Sigma^{-1/2} b = \lambda_{\max} b, \tag{7.1.17}$$

其中 $\lambda_{\max}$ 是最大特征值, b 是最大特征值 $\lambda_{\max}$ 所对应的特征向量,$\Sigma^{-1/2}=\left(\Sigma^{1/2}\right)^{-1}$, $\Sigma^{1/2}$ 是一个对称矩阵, 满足$\Sigma^{1/2}\Sigma^{1/2}=\Sigma$, 且 Σ 是 $\left(X^{\mathrm{T}}, Z^{\mathrm{T}}\right)^{\mathrm{T}}$ 的协方差矩阵. M 是 Zhu 等 (2010) 提出的累积分布降维矩阵:

$$M = E_{Y^*}\left[E_{X,Z,Y}\left\{\begin{pmatrix} X \\ Z \end{pmatrix} I\{Y \leqslant Y^*\}\right\} E_{X,Z,Y}\left\{\begin{pmatrix} X \\ Z \end{pmatrix} I\{Y \leqslant Y^*\}\right\}^{\mathrm{T}}\right],$$

其中 $E_A(\cdot)$ 表示对随机变量 A 或者随机向量 A 取数学期望, Y^* 是与 Y 独立同分布的随机变量. 进一步, Zhang 等 (2012c) 将上述步骤 1~步骤 3 的降维估计方法拓展到协变量维数发散情形下的参数估计与变量选择问题的研究, 读者可参考 Zhang 等 (2012c).

7.2 单指标测量误差模型的 SIMEX 估计

本节把第 1 章介绍的 SIMEX 方法应用到单指标测量误差模型, 对模型中的指标参数向量和联系函数进行估计. 在一定的正则条件下证明了指标参数向量的 SIMEX 估计的渐近正态性, 并推导了联系函数估计量的渐近偏差和渐近方差.

7.2.1 模型介绍

单指标模型可以克服多元非参数回归模型拟合的 "维数祸根" 问题, 是一个重要的半参数模型. 本节考虑如下的单指标测量误差模型:

$$\begin{cases} Y = g(\beta^{\mathrm{T}} Z) + \varepsilon, \\ W = Z + U, \end{cases} \tag{7.2.1}$$

其中 $g(\cdot)$ 是一个一元的未知联系函数, $\beta = (\beta_1, \cdots, \beta_p)^{\mathrm{T}}$ 是 $p \times 1$ 的未知指标参数向量, Y 是响应变量, ε 是随机的模型误差, 满足 $E(\varepsilon|Z) = 0$. 模型 (7.2.1) 中 p 维协变量 X 不能被准确测量, 而测量到的是具有测量误差 U 的协变量 W, 假设测量误差 $U \sim N(0, \Sigma_{uu})$, 且独立于 (Z, Y). 如果 U 的一些分量等于 0 时, 意味着对应的协变量分量没有测量误差. 由于联系函数 $g(\cdot)$ 未知, 考虑到模型的可识别性, 假设 $\|\beta\| = 1$. 模型 (7.2.1) 称为单指标测量误差模型或单指标 EV 模型, 本节仅考虑测量误差协方差阵 Σ_{uu} 已知的情形. 如果 Σ_{uu} 未知时, 可采用重复测量数据进行相合估计 Σ_{uu}, 前面章节已有讨论, 本节不再讨论.

对于单指标测量误差的研究, Liang 和 Wang (2005) 研究了部分线性单指标测量误差模型, 他们仅仅考虑线性部分带有测量误差的情形, 没有考虑非参数部分带有测量误差的情形. Liang 和 Wang (2005) 利用纠偏衰减的方法得到了指标参数分量的有效估计, 并研究了估计量的渐近性质. 但是如果非参数分量带有测量误差时, Liang 和 Wang (2005) 提出的估计方法不再适用, 这时要发展新的估计方法.

下面介绍 SIMEX 方法在单指标测量误差模型 (7.2.1) 中的应用.

7.2.2 SIMEX 估计程序

本节把 Cook 和 Stefanski (1994) 提出的 SIMEX 方法用于单指标测量误差模型 (7.2.1), 结合非参数的局部线性估计方法和估计方程, 对模型中的未知指标参数 β 和联系函数 $g(\cdot)$ 进行估计. 假设 $\{(Y_i, W_i), i = 1, \cdots, n\}$ 是来自模型 (7.2.1) 的 i.i.d. 随机样本, 下面给出估计程序.

(1) **模拟步** 对于每一个 $i = 1, \cdots, n$, 产生随机变量序列

$$W_{ib}(\lambda) = W_i + (\lambda \Sigma_{uu})^{1/2} U_{ib}, \quad b = 1, \cdots, B,$$

其中 $U_{ib} \sim N(0, I_p)$, I_p 是 $p \times p$ 的单位阵, B 是一个给定的任意大的正整数, $\lambda \in \Lambda = \{\lambda_1, \lambda_2, \cdots, \lambda_M\}$, 这里 Λ 表示 λ 取值的格子点, 一般 λ 的取值范围取为 [0,2].

(2) **估计步** 假设联系函数 $g(\cdot)$ 有连续的二阶导数, 对于 t_0 邻域内的点 t, $g(t)$ 能够用一个线性函数进行逼近, 即

$$g(t) \approx g(t_0) + g'(t_0)(t - t_0) =: a + b(t - t_0).$$

下面介绍基于数据集 $\{(Y_i, W_{ib}(\lambda)), i = 1, \cdots, n, b = 1, \cdots, B\}$, 关于未知指标参数 β 和联系函数 $g(\cdot)$ 的估计问题.

步骤 1 对于给定的 t_0 和 β. 定义下面的加权最小二乘目标函数

$$\sum_{i=1}^{n} \left\{Y_i - a - b[\beta^{\mathrm{T}} W_{ib}(\lambda) - t_0]\right\}^2 K_h(\beta^{\mathrm{T}} W_{ib}(\lambda) - t_0), \tag{7.2.2}$$

其中 $K_h(\cdot) = h^{-1}K(\cdot/h)$, $K(\cdot)$ 是核函数, h 是窗宽. 关于 a 和 b, 极小化加权最小二乘目标函数 (7.2.2), 可得 a 和 b 的估计, 分别记为 $\hat{a}$ 和 $\hat{b}$, 则有 $\hat{g}(\beta, \lambda; t_0) = \hat{a}$ 和 $\hat{g}'(\beta, \lambda; t_0) = \hat{b}$. 令

$$\mathcal{M}_{ni}(\beta, \lambda; t_0) = \mathcal{U}_{ni}(\beta, \lambda; t_0) \Big/ \sum_{j=1}^{n} \mathcal{U}_{nj}(\beta, \lambda; t_0)$$

和

$$\widetilde{\mathcal{M}}_{ni}(\beta, \lambda; t_0) = \widetilde{\mathcal{U}}_{ni}(\beta, \lambda; t_0) \Big/ \sum_{j=1}^{n} \mathcal{U}_{nj}(\beta, \lambda; t_0),$$

其中

$$\mathcal{U}_{ni}(\beta, \lambda; t_0) = K_h(\beta^{\mathrm{T}} W_{ib}(\lambda) - t_0)\{S_{n,2}(\beta, \lambda; t_0) - [\beta^{\mathrm{T}} W_{ib}(\lambda) - t_0] S_{n,1}(\beta, \lambda; t_0)\},$$

$$\widetilde{\mathcal{U}}_{ni}(\beta, \lambda; t_0) = K_h(\beta^{\mathrm{T}} W_{ib}(\lambda) - t_0)\{[\beta^{\mathrm{T}} W_{ib}(\lambda) - t] S_{n,0}(\beta, \lambda; t_0) - S_{n,1}(\beta, \lambda; t_0)\}$$

和

$$S_{n,l}(\beta, \lambda; t_0) = \frac{1}{n} \sum_{i=1}^{n} (\beta^{\mathrm{T}} W_{ib}(\lambda) - t_0)^l K_h(\beta^{\mathrm{T}} W_{ib}(\lambda) - t_0), \quad l = 0, 1, 2.$$

简单计算, 可得

$$\hat{g}(\beta, \lambda; t_0) = \sum_{i=1}^{n} \mathcal{M}_{ni}(\beta, \lambda; t_0) Y_i, \tag{7.2.3}$$

$$\hat{g}'(\beta, \lambda; t_0) = \sum_{i=1}^{n} \widetilde{\mathcal{M}}_{ni}(\beta, \lambda; t_0) Y_i. \tag{7.2.4}$$

Chang 等 (2010) 与 Pang 和 Xue (2012) 指出, 如果使用相同的窗宽估计联系函数 $g(t)$ 和其导数 $g'(t)$ 时, $g'(t)$ 的估计比 $g(t)$ 的估计有慢的收敛速度. 因此, 引入另一个窗宽 h_1 来控制 $g'(t)$ 估计的变差, 即在估计 $g'(t)$ 时使用窗宽 h_1, 把式 (7.2.4) 定义的估计记为 $\hat{g}'_{h_1}(\beta,\lambda;t_0)$.

步骤 2 注意可识别性约束条件 $\|\beta\|=1$, 该约束意味着 β 的真实参数值在单位超球面的边界上, 这时 $g(\beta^{\mathrm{T}}Z_i)$ 关于 β 的偏导数不存在, 而在构造估计方程时, 需要 $g(\beta^{\mathrm{T}}Z_i)$ 关于 β 的偏导数. 为了解决这个问题, 可以采用 "去一分量" 方法 (Zhu and Xue, 2006) 转换 $\mathbb{R}^p$ 空间中单位超球面上的点到 $\mathbb{R}^p$ 空间单位球的内点. 具体思路是: 令 $\beta=(\beta_1,\cdots,\beta_p)^{\mathrm{T}}$, $\beta^{(r)}=(\beta_1,\cdots,\beta_{r-1},\beta_{r+1},\cdots,\beta_p)^{\mathrm{T}}$ 表示去掉 β 的第 r 个分量 β_r 以后的 $p-1$ 维参数向量. 为了简单, 不妨假定 β 的第 r 个分量 β_r 是一个正的分量; 否则考虑 $\beta_r=-(1-\|\beta^{(r)}\|^2)^{1/2}$. 这时有

$$\beta=\left(\beta_1,\cdots,\beta_{r-1},(1-\|\beta^{(r)}\|)^2)^{1/2},\beta_{r+1},\cdots,\beta_p\right)^{\mathrm{T}}. \tag{7.2.5}$$

真参数 $\beta^{(r)}$ 满足 $\|\beta^{(r)}\|<1$. 这时, β 在真参数 $\beta^{(r)}$ 的某个邻域内有限可微, 其 Jacobian 矩阵是

$$J_{\beta^{(r)}}=(\gamma_1,\cdots,\gamma_p)^{\mathrm{T}},$$

其中 $\gamma_s(1\leqslant s\leqslant p, s\neq r)$ 是一个 $(p-1)$ 维的单位向量, 当 $s<r$ 时, 其第 s 个元素为 1, 当 $s>r$ 时, 其第 $s-1$ 个元素为 1, 而 $\gamma_r=-(1-\|\beta^{(r)}\|^2)^{-1/2}\beta^{(r)}$. 考虑分别定义在式 (7.2.3) 和式 (7.2.4) 的估计 $\hat{g}(\beta,\lambda;t_0)$ 和 $\hat{g}'_{h_1}(\beta,\lambda;t_0)$, 则 $\beta^{(r)}$ 的估计 $\hat{\beta}_b^{(r)}(\lambda)$ 可以解下面的估计方程得到

$$\frac{1}{n}\sum_{i=1}^{n}\hat{\eta}_{ib}(\beta,\lambda)=0, \tag{7.2.6}$$

其中

$$\hat{\eta}_{ib}(\beta,\lambda)=[Y_i-\hat{g}(\beta,\lambda;\beta^{\mathrm{T}}W_{ib}(\lambda))]\hat{g}'_{h_1}(\beta,\lambda;\beta^{\mathrm{T}}W_{ib}(\lambda))J^{\mathrm{T}}_{\beta^{(r)}}W_{ib}(\lambda).$$

因此, 由式 (7.2.5) 和估计 $\hat{\beta}_b^{(r)}(\lambda)$, 通过转化可得指标参数 β 的估计, 记为 $\hat{\beta}_b(\lambda)$. 为了解方程 (7.2.6), 需要使用 Newton-Raphson 迭代算法.

步骤 3 对 $\hat{\beta}_b(\lambda)$ 关于 $b=1,\cdots,B$ 进行平均, 则 $\hat{\beta}(\lambda)$ 定义为

$$\hat{\beta}(\lambda)=\frac{1}{B}\sum_{b=1}^{B}\hat{\beta}_b(\lambda).$$

(3) **外推步** 定义外推函数 $\mathcal{G}(\lambda,\Gamma)$. 基于外推函数 $\mathcal{G}(\lambda,\Gamma)$, 对 $\{\hat{\beta}(\lambda),\lambda\in\Lambda\}$ 在 $\{\lambda\in\Lambda\}$ 上关于 Γ 作拟合回归模型, 可以获得 Γ 的估计, 记为 $\hat{\Gamma}$. 最后可得到 β 的

SIMEX 估计定义为

$$\hat{\beta}_{\text{SIMEX}} = \mathcal{G}(-1, \hat{\Gamma}).$$

注意到: 如果 λ 退化到 0, $\hat{\beta}_{\text{Naive}} = \mathcal{G}(0, \hat{\Gamma})$ 就表示忽略测量误差后得到的估计, 即直接用 W 代替 Z 得到的估计量.

联系函数 $g(\cdot)$ 的 SIMEX 估计也可以用相同的方式得到. 在估计步中的步骤 1 中, 用 SIMEX 估计 $\hat{\beta}_{\text{SIMEX}}$ 代替 β, 并且用窗宽 h_2 获得估计 $\hat{g}_b(\lambda; t_0)$, 然后关于 $b = 1, \cdots, B$ 作平均, 则有

$$\hat{g}(\lambda; t_0) = \frac{1}{B} \sum_{b=1}^{B} \hat{g}_b(\lambda; t_0).$$

关于 $\mathbb{A}$, 极小化下式

$$\sum_{\lambda \in \Lambda} \{\hat{g}(\lambda; t_0) - \mathcal{G}(\lambda; \mathbb{A})\}^2,$$

则可获得估计 $\hat{\mathbb{A}}$. 进一步, 可获得联系函数 $g(\cdot)$ 的 SIMEX 估计为

$$\hat{g}_{\text{SIMEX}}(t_0) = \mathcal{G}(-1, \hat{\mathbb{A}}).$$

7.2.3 渐近性质

为了研究指标参数和联系函数的 SIMEX 估计的渐近性质, 首先列出下面的正则条件.

(C7.2.1) 假设 $\beta^{\text{T}} Z$ 的密度函数 $f(t)$ 是有界远离 0 的, 且在 $\mathcal{T} = \{t = \beta^{\text{T}} z : z \in A\}$ 上满足 1 阶 Lipschitz 条件, 其中 A 是 Z 的有界支撑集.

(C7.2.2) 联系函数 $g(\cdot)$ 在 $\mathcal{T}$ 有连续的二阶导数.

(C7.2.3) 核函数 $K(\cdot)$ 是有界对称的概率密度函数, 并满足

$$\int_{-\infty}^{\infty} u^2 K(u) \text{d}u \neq 0, \quad \int_{-\infty}^{\infty} |u|^i K(u) \text{d}u < \infty, \quad i = 1, 2, \cdots.$$

(C7.2.4) 假设 $\sup\limits_z E(\varepsilon^2 | Z = z) < \infty$, $\sup\limits_z E(\varepsilon^4 | Z = z) < \infty$.

(C7.2.5) 当 $n \to \infty$ 时, $nh^2/\log^2 n \to \infty, nh^4 \to 0$, $nhh_1^3/\log^2 n \to \infty$, $\lim\sup\limits_{n \to \infty} nh_1^5 < \infty$.

(C7.2.6) 对于 $\lambda \in \Lambda$, 假设 $\mathcal{A}(\beta(\lambda), \lambda)$ 是一个正定矩阵, 其中

$$\mathcal{A}(\beta(\lambda), \lambda) = E\left\{ \left[g'\left(\lambda; \beta^{\text{T}}(\lambda) W_{ib}(\lambda)\right) \right]^2 J_{\beta^{(r)}(\lambda)}^{\text{T}} \widetilde{W}_{ib}(\lambda) \widetilde{W}_{ib}^{\text{T}}(\lambda) J_{\beta^{(r)}(\lambda)} \right\},$$

且 $\widetilde{W}_{ib}(\lambda) = W_{ib}(\lambda) - E[W_{ib}(\lambda) | \beta^{\text{T}}(\lambda) W_{ib}(\lambda)]$.

(C7.2.7) 外推函数 $\mathcal{G}(\lambda, \Gamma)$ 在理论上是精确的.

为了给出下面的渐近结果, 首先定义一些概念. 定义 $\mathcal{G}(\Lambda,\mathbf{\Gamma})=\mathrm{vec}\{\mathcal{G}(\lambda_m,\Gamma_j), j=1,\cdots,p,m=1,\cdots,M\}$, 其中 $\mathbf{\Gamma}=(\Gamma_1^{\mathrm{T}},\cdots,\Gamma_p^{\mathrm{T}})^{\mathrm{T}}$, Γ_j 为外推步中对第 j 个分量估计 $\hat{\beta}(\lambda)$ 时, 外推函数中的参数向量, $j=1,\cdots,p$. 此处为了区别, 用黑体表示 $\mathbf{\Gamma}$. 进一步, 定义 $\mathrm{Res}(\mathbf{\Gamma})=\hat{\beta}(\Lambda)-\mathcal{G}(\Lambda,\mathbf{\Gamma})$. $\hat{\mathbf{\Gamma}}$ 是在外推步中通过极小化 $\{\mathrm{Res}(\mathbf{\Gamma})\}\{\mathrm{Res}(\mathbf{\Gamma})\}^{\mathrm{T}}$ 所得估计, 即 $\hat{\mathbf{\Gamma}}$ 是估计方程 $s(\mathbf{\Gamma})\mathrm{Res}(\mathbf{\Gamma})=0$ 的解, 其中 $s^{\mathrm{T}}(\mathbf{\Gamma})=\{\partial/\partial(\mathbf{\Gamma})^{\mathrm{T}}\}\mathrm{Res}(\mathbf{\Gamma})$. 令

$$\eta_{iB}(\beta(\lambda),\lambda)=\frac{1}{B}\sum_{b=1}^{B}\left[Y_i-g\Big(\lambda;\beta^{\mathrm{T}}(\lambda)W_{ib}(\lambda)\Big)\right]g'\Big(\lambda;\beta^{\mathrm{T}}(\lambda)W_{ib}(\lambda)\Big)J_{\beta^{(r)}(\lambda)}^{\mathrm{T}}\widetilde{W}_{ib}(\lambda)$$

和 $D(\mathbf{\Gamma})=s(\mathbf{\Gamma})s^{\mathrm{T}}(\mathbf{\Gamma})$, 则可以证明

$$\sqrt{n}(\hat{\mathbf{\Gamma}}-\mathbf{\Gamma})\xrightarrow{L}N\{0,\Sigma(\mathbf{\Gamma})\},$$

其中 $\Sigma(\mathbf{\Gamma})=D^{-1}(\mathbf{\Gamma})s(\mathbf{\Gamma})\Sigma s^{\mathrm{T}}(\mathbf{\Gamma})D^{-1}(\mathbf{\Gamma})$, 这里 Σ 的定义为

$$\Sigma=\mathcal{J}\Big(\beta(\Lambda),\Lambda\Big)\mathcal{A}_{11}^{-1}\Big(\beta(\Lambda),\Lambda\Big)C_{11}\Big(\beta(\Lambda),\Lambda\Big)\Big\{\mathcal{A}_{11}^{-1}\Big(\beta(\Lambda),\Lambda\Big)\Big\}^{\mathrm{T}}\mathcal{J}^{\mathrm{T}}\Big(\beta(\Lambda),\Lambda\Big),$$

其中

$$\mathcal{J}\Big(\beta(\Lambda),\Lambda\Big)=\mathrm{diag}\{J_{\beta^{(r)}(\lambda)},\lambda\in\Lambda\},\quad \mathcal{A}_{11}\Big(\beta(\Lambda),\Lambda\Big)=\mathrm{diag}\{\mathcal{A}(\beta(\lambda),\lambda),\lambda\in\Lambda\}$$

和

$$C_{11}\Big(\beta(\Lambda),\Lambda\Big)=\mathrm{Cov}\Big(\mathrm{vec}\{\eta_{iB}(\beta(\lambda),\lambda),\lambda\in\Lambda\}\Big).$$

定理 7.2.1　*假设正则条件(C7.2.1)～条件(C7.2.7)成立. 当 $n\to\infty$ 时, 则有 $\sqrt{n}(\hat{\beta}_{\mathrm{SIMEX}}-\beta)$ 是渐近正态的, 均值为 0, 渐近协方差阵由式 (7.2.7) 所定义*

$$\mathcal{G}_{\mathbf{\Gamma}}(-1,\mathbf{\Gamma})\Sigma(\mathbf{\Gamma})\{\mathcal{G}_{\mathbf{\Gamma}}(-1,\mathbf{\Gamma})\}^{\mathrm{T}},\tag{7.2.7}$$

其中 $\mathcal{G}_{\mathbf{\Gamma}}(\lambda,\mathbf{\Gamma})=\{\partial/\partial(\mathbf{\Gamma})^{\mathrm{T}}\}\mathcal{G}(\lambda,\mathbf{\Gamma})$.

定理 7.2.1 暗示着 $\hat{\beta}_{\mathrm{SIMEX}}$ 是一个 $\sqrt{n}$ 相合估计, 并且 $\hat{\beta}_{\mathrm{SIMEX}}$ 的渐近分布是类似于没有测量误差时参数估计的渐近分布, 只是渐近协方差阵变得更加复杂, 因为测量误差会引起协方差阵发生改变.

令 $f_0(\cdot)$ 是 $\beta^{\mathrm{T}}W$ 的密度函数, $\mu_l=\int t^lK(t)\mathrm{d}t$ 和 $\nu_l=\int K^l(t)\mathrm{d}t$, $l=1,2$. 定义

$$\gamma(\lambda,\mathbb{A})=\{\partial/\partial(\mathbb{A})\}\mathcal{G}(\lambda,\mathbb{A}),\quad C(\Lambda,\mathbb{A})=\gamma^{\mathrm{T}}(-1,\mathbb{A})\Big\{\sum_{\lambda\in\Lambda}\gamma(\lambda,\mathbb{A})\gamma^{\mathrm{T}}(\lambda,\mathbb{A})\Big\}^{-1}$$

和

$$D = E_q\gamma(\lambda, \mathbb{A})\gamma^{\mathrm{T}}(\lambda, \mathbb{A})E_q,$$

其中 E_q 是 $q\times q$ 的矩阵, 该矩阵第 1 个元素为 1, 其他所有元素都为 0, 且 q 表示 $\mathbb{A}$ 的维数.

定理 7.2.2 假设正则条件(C7.2.1)~条件(C7.2.7)成立. 如果 $nh_2^5 = O(1)$, 当 $n\to\infty$ 和 $B\to\infty$ 时, 则联系函数的 SIMEX 估计 $\hat{g}_{\mathrm{SIMEX}}(t_0)$ 的渐近偏差和渐近方差分别为

$$C(\Lambda, \mathbb{A})\sum_{\lambda\in\Lambda}\frac{1}{2}h_2^2\mu_2 g''(\lambda; t_0)\gamma(\lambda, \mathbb{A})$$

和

$$[nh_2 f_0(t_0)]^{-1}\nu_2\mathrm{Var}(Y|\beta^{\mathrm{T}}W = t_0)C(\Lambda, \mathbb{A})DC^{\mathrm{T}}(\Lambda, \mathbb{A}),$$

其中 $g(\lambda; t) = E(Y|\beta^{\mathrm{T}}W_b(\lambda) = t)$.

定理 7.2.2 说明指标参数的 SIMEX 估计 $\hat{\beta}_{\mathrm{SIMEX}}$ 并不影响联系函数的 SIMEX 估计 $\hat{g}_{\mathrm{SIMEX}}(t_0)$, 因为 SIMEX 估计 $\hat{\beta}_{\mathrm{SIMEX}}$ 是 $\sqrt{n}$ 相合的. Carroll 等 (1999) 指出, 如果忽略测量误差时, $\hat{g}_{\mathrm{SIMEX}}(t_0)$ 的渐近方差渐近等价于不存在测量误差时联系函数的渐近方差乘以因子 $C(\Lambda, \mathbb{A})DC^{\mathrm{T}}(\Lambda, \mathbb{A})$.

7.2.4 模拟研究和实际数据分析

1. 模拟研究

本节通过模拟研究来评价 SIMEX 方法在有限样本下的完成情况, 考虑下面的单指标测量误差模型:

$$\begin{cases} Y_i = -2(\beta^{\mathrm{T}}Z_i - 1)^2 + 1 + \varepsilon_i, \\ W_i = Z_i + U_i, \quad i = 1, \cdots, n, \end{cases}$$

其中 $\beta = (\beta_1, \beta_2)^{\mathrm{T}} = (\sqrt{3}/3, \sqrt{6}/3)^{\mathrm{T}}$, Z_i 是一个二维的随机向量, 每个分量独立的来自标准正态分布 $N(0, 1)$, 模型误差 $\varepsilon_i \sim N(0, 0.2^2)$, 响应变量 Y_i 从第一个模型中产生. 测量误差 $U_i \sim N(0, \mathrm{diag}(\sigma_u^2, 0))$, 取 $\sigma_u = 0.2, 0.4, 0.6$ 表示不同的测量误差水平. 样本量大小取 $n = 50, 100$ 和 150. 对于每种情形, 重复模拟 500 次对所提方法进行评价. 对于 SIMEX 算法, 取 $\lambda = 0, 0.2, \cdots, 2$ 和 $B = 50$. 对于联系函数的估计中, 使用 Epanechnikov 核函数 $K(u) = 0.75(1-u^2)I(|u|\leqslant 1)$. 关于光滑参数窗宽的选取, 尽管可以采用交错核实方法选取窗宽, 但是为了节省计算时间, 窗宽 h,h_1 和 h_2 可分别取为 $cn^{-1/4}(\log n)^{-1/2}$, $cn^{-1/5}$ 和 $cn^{-1/5}$, 其中 c 为 $\beta_{\mathrm{int}}^{\mathrm{T}}Z$ 的标准差, 且 β_{int} 为在约束条件 $\|\beta\| = 1$ 下, 通过拟合一个线性模型得到的初始值.

在模拟研究中, 比较了忽略测量误差的自然估计 (Naive) 和具有二次外推函数的 SIMEX 估计, 表 7.2.1 中提供了两种估计的偏差 (Bias) 和标准差 (SD), 通过偏差和标准差比较两种估计完成的好坏. 从表 7.2.1 能够看出, β_1 和 β_2 的 SIMEX 估计比自然估计有小的偏差, 然而基于 SIMEX 估计的标准差却比自然估计的要大. 另外也可以发现, 随着样本量 n 的变大, 两种估计的偏差和标准差都变小, 并且两种估计的结果都依赖于测量误差的大小.

表 7.2.1　β_1 和 β_2 的 SIMEX 估计和自然估计的偏差 (Bias) 和标准差 (SD)

n	σ_u	SIMEX		Naive	
		β_1	β_2	β_1	β_2
		Bias(SD)	Bias(SD)	Bias(SD)	Bias(SD)
50	0.2	−0.0084(0.0520)	0.0078(0.0377)	−0.0177(0.0291)	0.0146(0.0203)
	0.4	−0.0405(0.0875)	0.0171(0.0638)	−0.0764(0.0537)	0.0546(0.0388)
	0.6	−0.0508(0.1253)	0.0342(0.0821)	−0.1207(0.0680)	0.0700(0.0330)
100	0.2	−0.0083(0.0384)	0.0074(0.0321)	−0.0126(0.0203)	0.0084(0.0142)
	0.4	−0.0381(0.0581)	0.0158(0.0334)	−0.0761(0.0397)	0.0434(0.0224)
	0.6	−0.0394(0.0719)	0.0206(0.0567)	−0.1154(0.0383)	0.0632(0.0210)
150	0.2	−0.0059(0.0169)	0.0039(0.0118)	−0.0187(0.0136)	0.0127(0.0093)
	0.4	−0.0160(0.0341)	0.0126(0.0258)	−0.0497(0.0279)	0.0324(0.0177)
	0.6	−0.0279(0.0599)	0.0163(0.0394)	−0.1088(0.0315)	0.0563(0.0171)

为了讨论联系函数 $g(t)$ 估计 $\hat{g}(t)$ 的完成情况, 用下面定义的均方根误差 (root mean squared error, RMSE) 进行评价

$$\mathrm{RMSE} = \left[n_{\mathrm{grid}}^{-1} \sum_{k=1}^{n_{\mathrm{grid}}} \{\hat{g}(t_k) - g(t_k)\}^2 \right]^{1/2},$$

其中 n_{grid} 是格子点数, 且 $\{t_k, k = 1, 2, \cdots, n_{\mathrm{grid}}\}$ 是等间距的格子点, 模拟计算中取 $n_{\mathrm{grid}} = 15$. 图 7.2.1 展示了基于 500 次重复计算所得联系函数平均估计的拟合曲线图和 RMSE 的箱线图. 从图 7.2.1(a), 可以看到联系函数的 SIMEX 估计的拟合曲线接近于真实联系函数曲线, 而忽略测量误差得到的自然估计曲线与真实曲线有一定的偏差. 图 7.2.1(b) 展示了联系函数 $g(\cdot)$ 的 SIMEX 估计和自然估计的均方根误差 (RMSE) 的箱线图, 发现 SIMEX 估计的 RMSE 和自然估计的 RMSE 差不多, 但是 SIMEX 估计的 RMSE 比自然估计的 RMSE 要轻微大一点.

从模拟结果可以看出, SIMEX 方法有效地减少了估计量的偏差, 说明是测量误差模型纠偏的一个有效方法.

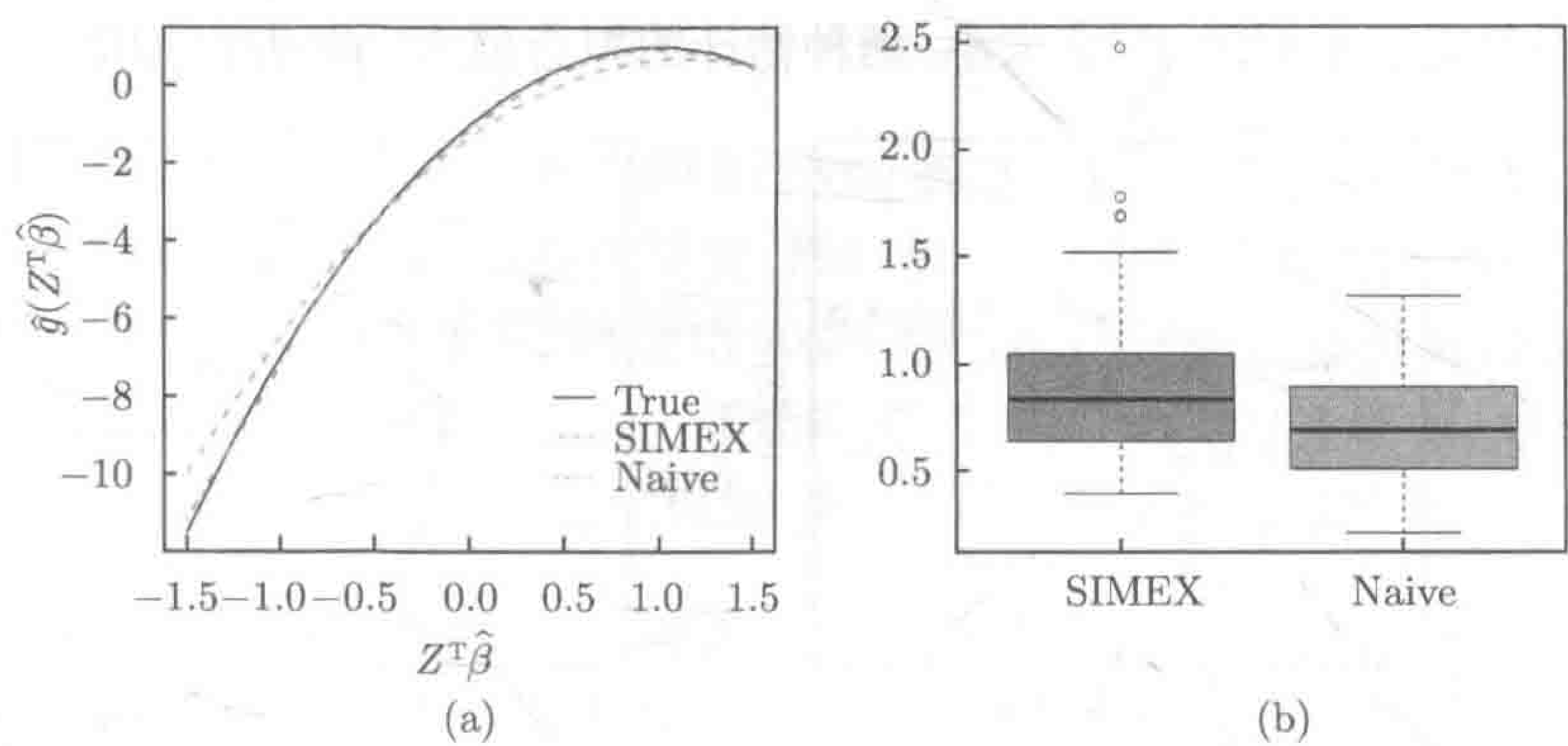


图 7.2.1 (a) 实线表示联系函数 $g(t)$ 的真实曲线, 断线表示自然估计的拟合曲线, 点断线表示 SIMEX 估计的拟合曲线, 其中 $n = 100$ 和 $\sigma_u = 0.4$; (b) 基于 500 次重复试验的联系函数 $g(t)$ 两种估计的箱线图

2. 实际数据分析

下面对一个心脏疾病实际数据进行分析来说明所提出的 SIMEX 方法. 数据来源于网站 http://mlr.cs.umass.edu/ml/datasets/Statlog+%28Heart%29, 该心脏病实际数据由 270 个患者的 13 个变量构成的一个数据集. 感兴趣的问题是研究是否安静血压和血清胆固醇对最大心率有显著性的影响. 对该数据集用单指标测量误差模型进行建模和分析, 并比较 SIMEX 估计和自然估计两种方法. 令 Y 表示最大心率, W_1 表示安静血压的对数, 即 $\log(BP)$, W_2 表示血清胆固醇的对数, 即 $\log(SC)$, 其中 BP 表示安静血压, SC 表示血清胆固醇. 在实际研究中, 这两个变量很难准确测量, 存在一定的测量误差. 下面先作一个敏感性分析, 假设 Σ_{uu} 是已知的, 取 $\Sigma_{uu} = 0.4^2 I_2$ 和 0 矩阵. 表 7.2.2, 图 7.2.2 和图 7.2.3 提供了计算结果.

表 7.2.2 对心脏疾病数据, 指标参数的 SIMEX 估计和 Naive 估计

方法	$\log(BP)$	$\log(SC)$
SIMEX	0.9023	0.4297
Naive	0.9040	0.4274

从表 7.2.2, 发现变量 $\log(BP)$ 的指标系数的 SIMEX 估计比自然估计要小, 而变量 $\log(SC)$ 的指标系数的 SIMEX 估计比自然估计要大. 该结果表明安静血压和血清胆固醇对最大心率都有显著性的影响. 图 7.2.2 展示了 SIMEX 估计在外推步的轨迹, 给出了对不同的 λ 值指标系数向量的估计, SIMEX 估计是横坐标为 $\lambda = -1$ 所对应的值, 而自然估计是横坐标为 $\lambda = 0$ 所对应的值. 图 7.2.3 给出了联系函数 $g(\cdot)$ 的 SIMEX 估计的拟合曲线和自然估计的拟合曲线. 这两个拟合曲线的趋势比

较类似, 但是对于不同的 $Z^{\mathrm{T}}\hat{\beta}$ 大小, 两种估计的拟合值大小是有区别的.

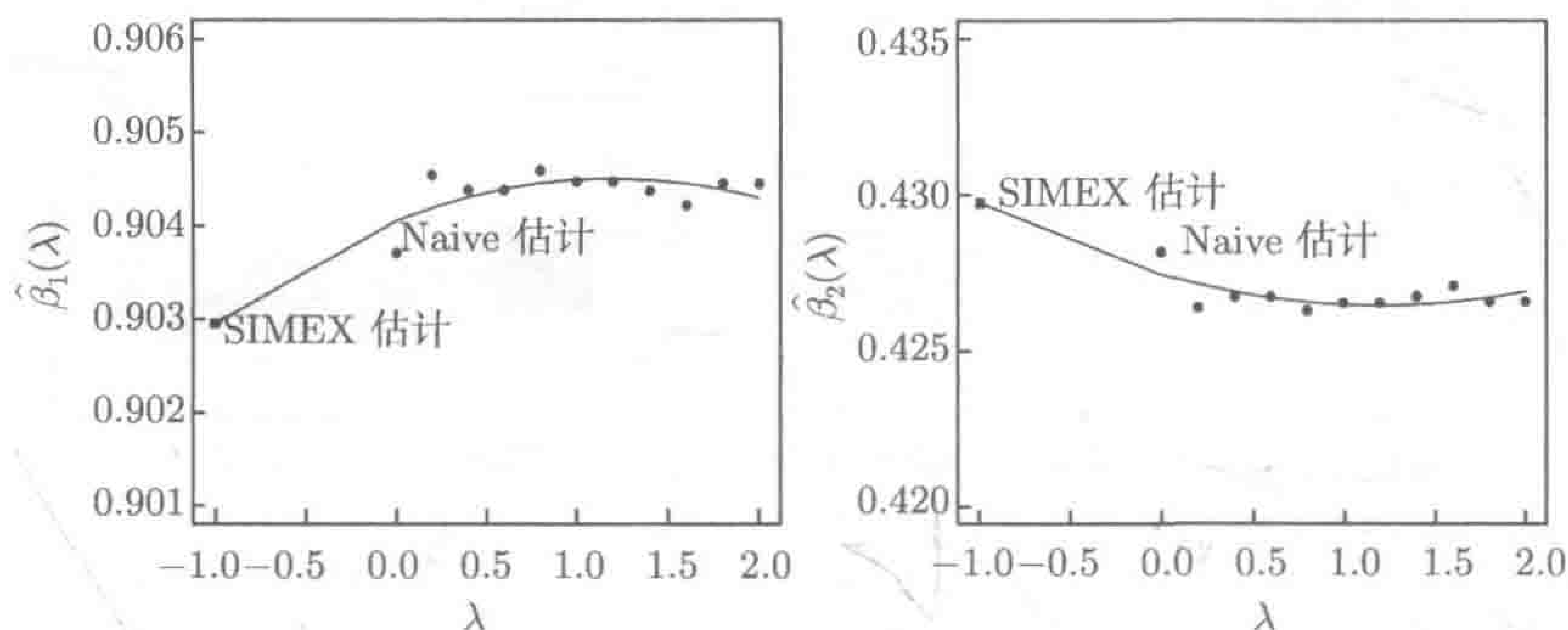


图 7.2.2　心脏疾病数据的外推点估计曲线, 图中的点表示模拟估计 $\{\hat{\beta}(\lambda),\lambda\}$, 实线表示基于二次外推函数的拟合曲线, 并且被外推到 $\lambda=-1$ 的情况

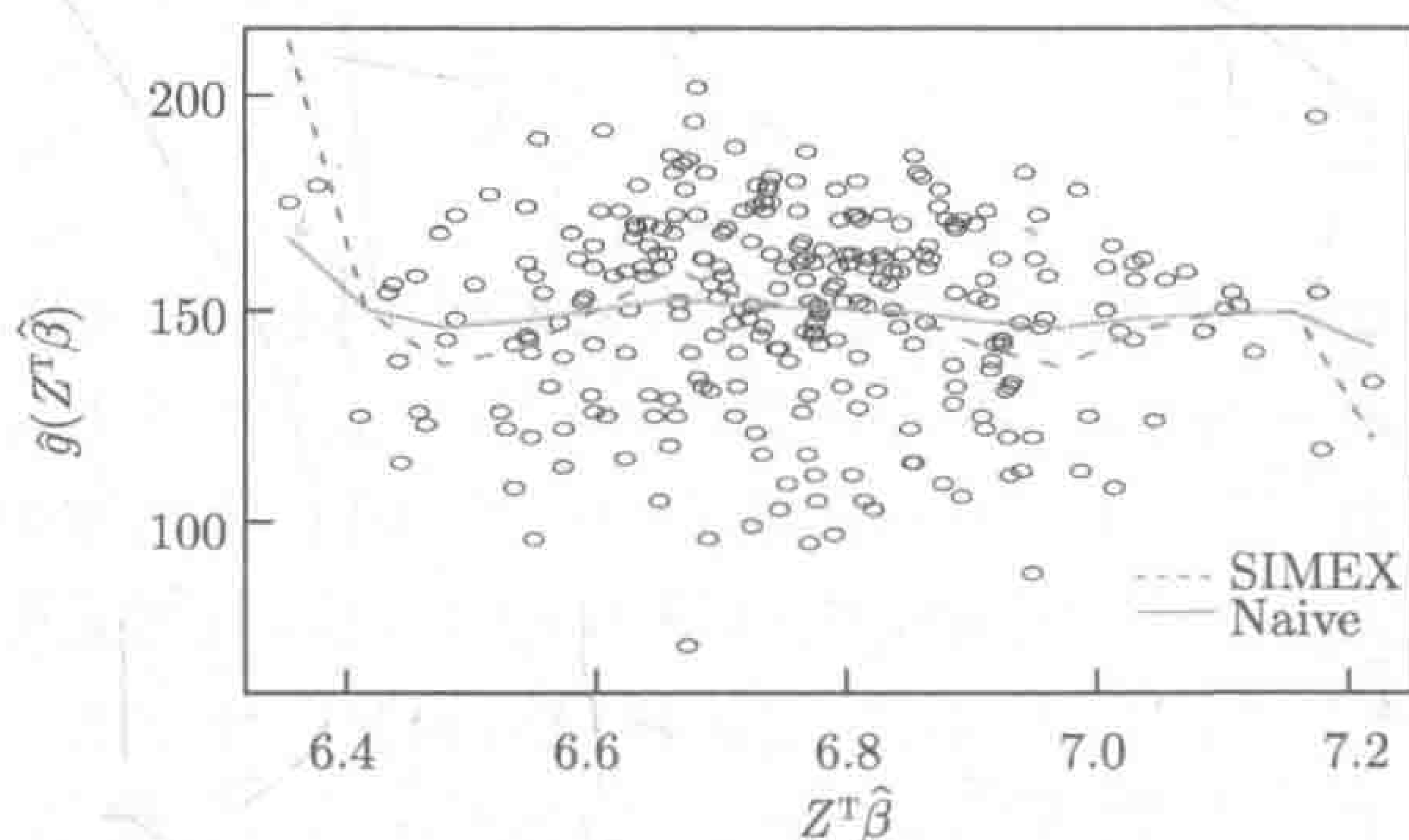


图 7.2.3　关于心脏疾病数据分析的联系函数的拟合曲线, 实线表示自然估计拟合曲线, 虚线表示 SIMEX 估计的拟合曲线

7.3　部分线性单指标 EV 模型

本节考虑协变量 Z 准确测量, 而协变量 X 具有可加测量误差的部分线性单指标 EV 模型:

$$\begin{cases} Y = X^{\mathrm{T}}\beta + g(Z^{\mathrm{T}}\theta) + \varepsilon, \\ W = X + U. \end{cases} \tag{7.3.1}$$

假设$\{(Y_i, W_i=(W_{i1},\cdots,W_{ip})^{\mathrm{T}}, Z_i=(Z_{i1},\cdots,Z_{iq})^{\mathrm{T}}), 1\leqslant i\leqslant n\}$ 是一组 i.i.d. 可观测随机样本. 在本节所介绍的三种估计方法中, 都假设 U 的协方差矩阵 Σ_{uu} 已知. 如果 Σ_{uu} 未知时, 可采用重复测量数据进行相合估计 Σ_{uu}, 前面章节已有讨论, 本节不再讨论.

7.3.1 两阶段估计方法

两阶段估计方法由 Liang 和 Wang (2005) 提出, 该估计方法的主要思想是将部分线性单指标模型转变为部分线性模型, 然后基于第 5 章介绍的部分线性 EV 模型的估计方法, 先给出 β 的估计, 再使用单指标模型的方法给出指标参数 θ 的估计. 这里先给出总体情形下的估计方法. 基于模型 (7.3.1), 给定协变量 Z, 取条件期望有 $E(Y|Z)=E(X|Z)^{\mathrm{T}}\beta+g(Z^{\mathrm{T}}\theta)$. 模型 (7.3.1) 的第一式和上式作差, 有

$$Y-E(Y|Z)=\{X-E(X|Z)\}^{\mathrm{T}}\beta+\varepsilon. \tag{7.3.2}$$

注意到 U 与 (Y,X,Z) 相互独立这一假设条件和 $E(W|Z)=E(X|Z)$ 成立, 如果直接用观测变量 W 代替 X, 结合第 5 章的部分线性 EV 模型结构下的估计方法, 可构造如下的总体最小二乘目标函数:

$$\begin{aligned}&E\left[Y-E(Y|Z)-\{W-E(W|Z)\}^{\mathrm{T}}\beta\right]^2\\=&E\left[Y-E(Y|Z)-\{X-E(X|Z)\}^{\mathrm{T}}\beta\right]^2+\beta^{\mathrm{T}}\Sigma_{uu}\beta.\end{aligned} \tag{7.3.3}$$

因此, 为了获得参数 β 的相合估计, Liang 和 Wang (2005) 构造了如下校正的最小二乘目标函数:

$$Q(\beta)=\frac{1}{n}\sum_{i=1}^{n}\left[Y_i-E(Y_i|Z_i)-\{W_i-E(W_i|Z_i)\}^{\mathrm{T}}\beta\right]^2-\beta^{\mathrm{T}}\Sigma_{uu}\beta.$$

在校正的最小二乘目标函数 $Q(\beta)$ 中包含未知的函数 $E(Y_i|Z_i)$ 和 $E(W_i|Z_i)$, 可利用核估计方法进行估计, 定义为

$$\hat{g}_W(z)=\sum_{i=1}^{n}\omega_{ni}(z)W_i,\quad \hat{g}_Y(z)=\sum_{i=1}^{n}\omega_{ni}(z)Y_i. \tag{7.3.4}$$

即 $\hat{g}_W(\cdot)$ 和 $\hat{g}_Y(\cdot)$ 分别是 $E(W|Z=\cdot)$ 和 $E(Y|Z=\cdot)$ 的核估计. Liang 和 Wang (2005) 选用的权函数为 $\omega_{ni}(z)=\omega_{ni}(z;Z_1,\cdots,Z_n)$, 其中 $\omega_{ni}(z)=n^{-1}K_H(Z_i-z)$, $K_H(z)=|H|^{-1/2}K_1(H^{-1/2}z)$. $K_1(\cdot)$ 是一个定义在有界紧集上的 q 维有界核密度函数, 满足 $\displaystyle\int K_1(z)\mathrm{d}z=1$. 这里窗宽矩阵 H 是一个依赖于样本 n 的 $q\times q$ 维的正定窗宽矩阵. Wand 和 Jones (1995) 讨论了多维核密度估计中最优窗宽矩阵 H 的一些选取方法. 为简单起见, Liang 和 Wang (2005) 将窗宽矩阵 H 指定为对角矩阵和 $K_1(z)=\displaystyle\prod_{s=1}^{q}K_{s,h_s}(z_s)$, 其中 $K_{s,h_s}(z_s)=h_s^{-1}K_s(z_s/h_s)$, 且 $K_s(\cdot)$ 是 1 维对称核密度函数.

把式 (7.3.4) 定义的核估计代入到校正的最小二乘目标函数 $Q(\beta)$ 中, 并关于未知参数 β 极小化, 则得到 β 的估计为

$$\hat{\beta}_n^* = \left[\sum_{i=1}^n \{W_i - \hat{g}_W(Z_i)\}\{W_i - \hat{g}_W(Z_i)\}^{\mathrm{T}} - n\Sigma_{uu}\right]^{-1} \times \sum_{i=1}^n \{W_i - \hat{g}_W(Z_i)\}\{Y_i - \hat{g}_Y(Z_i)\}. \tag{7.3.5}$$

为了得到估计量的渐近结果, 需要下面的假设条件.

(C7.3.1) 矩阵 $\Gamma_{X|Z} = E\left[(X - E(X|Z))^{\otimes 2}\right]$ 为正定矩阵.

(C7.3.2) 由函数 $E(X|Z=z)$ 和 $E(Y|Z=z)$ 定义的 Hessian 矩阵中的每一个元素都是连续且平方可积的, 这里函数 $g(z)$ 的 Hessian 矩阵的第 (i,j) 个元素定义为 $\dfrac{\partial^2 g(z)}{\partial z_i \partial z_j}$.

(C7.3.3) 窗宽矩阵 H 需满足矩阵 $H^{1/2}$ 的对角线元素的速度为 $O(h_1)$, 其中 $H^{1/2}$ 是对称矩阵且有 $H^{1/2}H^{1/2} = H$, $h_1 \in [C_1 n^{-1/(q+4)}, C_2 n^{-1/(q+4)}]$, $0 < C_1 < C_2 < \infty$, 其中 q 为随机变量 Z 的维数.

(C7.3.4) 权函数 $\omega_{ni}(\cdot)$ 满足: (a) $\max\limits_{1\leqslant i\leqslant n}\sum_{j=1}^n \omega_{ni}(Z_j) = O_P(1)$; (b) $\max\limits_{1\leqslant i,j\leqslant n} \omega_{ni}(Z_j) = O_P(b_n)$; (c) $\max\limits_{1\leqslant i\leqslant n}\sum_{j=1}^n \omega_{nj}(Z_i) I(\|Z_i - Z_j\| > c_n) = O_P(c_n)$, 其中 $b_n = n^{-1+q/(q+4)}$, $c_n = n^{-1/(q+4)}\log n$.

条件 (C7.3.2) 和 Liang 等 (1999) 定理 3.1 中的条件等价. 条件 (C7.3.3) 指出 Z 的每个分量有相同的窗宽速度. 条件 (C7.3.4) 中的核函数和 Liang 等 (1999) 中的假设条件 3.1 类似, 这里只是将条件拓展到非参数部分 Z 为 q 维 $(q > 1)$ 的情形.

定理 7.3.1 假设条件(C7.3.1)~条件(C7.3.4)和条件 $E(|\varepsilon|^4 + \|U\|^4) < \infty$ 成立, 那么当 $n \to \infty$ 时, $\hat{\beta}_n^*$ 是渐近正态的, 即

$$\sqrt{n}(\hat{\beta}_n^* - \beta) \xrightarrow{L} N\left(0, \Gamma_{X|Z}^{-1}\Sigma_\beta\Gamma_{X|Z}^{-1}\right),$$

其中 $\Sigma_\beta = E\left[(\varepsilon - U^{\mathrm{T}}\beta)^2\{X - E(X|Z)\}^{\otimes 2}\right] + E\left[\{U^{\otimes 2}\beta - \Sigma_{uu}^{\mathrm{T}}\beta\}^{\otimes 2}\right] + E(U^{\otimes 2}\varepsilon^2)$.

此外, 如果 ε 同方差并和 (X,Z) 相互独立, 矩阵 Σ_β 可简化为 $E(\varepsilon - U^{\mathrm{T}}\beta)^2\Gamma_{X|Z} + E\left[\left\{U^{\otimes 2}\beta - \Sigma_{uu}^{\mathrm{T}}\beta\right\}^{\otimes 2}\right] + \Sigma_{uu}\sigma^2$, $\sigma^2 = \mathrm{Var}(\varepsilon)$.

定理 7.3.1 主要的证明过程和 Liang 等 (1999) 证明定理 3.1 的过程类似, 其中

最重要的目标是为了得到如下渐近表达式

$$\sqrt{n}\left(\hat{\beta}_n^*-\beta\right)$$
$$=n^{-1/2}\Gamma_{X|Z}^{-1}\sum_{i=1}^{n}\left[\{X_i-E(X_i|Z_i)\}(\epsilon_i-U_i^{\mathrm{T}}\beta)-\left(U_i^{\otimes 2}-\Sigma_{uu}\right)\beta+U_i\varepsilon_i\right]+o_P(1).$$

注意到在适当的窗宽条件下, 定理 7.3.1 的结果表明 $\hat{\beta}_n^*$ 的渐近方差和 Z 的维数无关. 但是在实际估计过程中, 核估计 $\hat{g}_W(z)$ 和 $\hat{g}_Y(z)$ 在有限样本的情况下, 随着 Z 维数的增大, 该估计的实际表现会越来越差.

在得到 β 的估计后, Liang 和 Wang (2005) 将模型 (7.3.1) 转换为如下的单指标模型

$$Y_i-W_i^{\mathrm{T}}\hat{\beta}_n^*=g(Z_i^{\mathrm{T}}\theta)+\varepsilon^*.$$

然后利用单指标模型的估计方法给出指标参数 θ 和联系函数 $g(\cdot)$ 的估计. 在现有文献中, 关于单指标模型的估计问题, 已有相当多的估计方法来给出 θ 的 $\sqrt{n}$ 相合估计和 $g(\cdot)$ 的 $\sqrt{nh}$ 相合估计, 其中 h 为估计 $g(\cdot)$ 时使用的窗宽. 读者可参阅文献 Härdle 等 (1993), Friedman 和 Stuetzle (1981) 的投影追踪方法, Härdle 和 Stoker (1989) 的平均导数方法和 Li (1991) 提出的切片逆回归方法.

Liang 和 Wang (2005) 注意到 $g(u,\theta,\beta)=E(Y-X^{\mathrm{T}}\beta|Z^{\mathrm{T}}\theta=u)=E(Y-W^{\mathrm{T}}\beta|Z^{\mathrm{T}}\theta=u)$. 即从总体意义上讲, 测量误差 U 不影响指标参数 θ 与联系函数 $g(\cdot)$ 的估计. 因此, Liang 和 Wang (2005) 提出了基于数据 $\{(Y_i-W_i^{\mathrm{T}}\hat{\beta}_n^*,Z_i),i=1,\cdots,n\}$ 的非参数核估计迭代方法. 关于联系函数 $g(\cdot)$ 的估计, Liang 和 Wang (2005) 建议使用 Severini 和 Staniswalis (1994), Carroll 等 (1997) 所使用的局部线性光滑方法. $g(v)$ 在 u 的邻域内进行局部线性逼近, 即 $g(v)\approx g(u)+g'(u)(v-u)=:a+b(v-u)$, 其中 $a=g(u),b=g'(u)$. 下面给出指标参数 θ 与联系函数 $g(u)$ 估计的迭代算法.

步骤 1 将 $Y_i-W_i^{\mathrm{T}}\hat{\beta}_n^*$ 作为新的响应变量, 得到 θ 的一个初始相合估计 $\hat{\theta}_{(0)}$, 如可以利用 Li (1991) 提出的切片逆回归方法, 并正则化该初始估计: $\hat{\theta}=\hat{\theta}_{(0)}/\|\hat{\theta}_{(0)}\|$.

步骤 2 记 $\hat{\Lambda}_i=Z_i^{\mathrm{T}}\hat{\theta}$. 关于 (a,b) 极小化下面的局部加权最小二乘目标函数:

$$(\hat{a}(u),\hat{b}(u))=\arg\min_{a,b}\sum_{i=1}^{n}\left\{Y_i-W_i^{\mathrm{T}}\hat{\beta}_n^*-a-b(\hat{\Lambda}_i-u)\right\}^2K_{2h_2}(\hat{\Lambda}_i-u),$$

其中 $K_{2h_2}(\cdot)=h_2^{-1}K_2(\cdot/h_2)$, $K_2(\cdot)$ 为满足条件 (C7.3.7) 的核密度函数, h_2 为窗宽.

步骤 3 极小化下面的最小二乘目标函数来更新 θ 的估计:

$$\widetilde{\theta}=\arg\min_{\theta^*}\sum_{i=1}^{n}\left\{Y_i-W_i^{\mathrm{T}}\hat{\beta}_n^*-\hat{a}(Z_i^{\mathrm{T}}\theta^*)\right\}^2,$$

并正则化 $\hat{\theta}=\widetilde{\theta}/\|\widetilde{\theta}\|$. 在步骤 2 和步骤 3 之间进行重复迭代, 直到收敛, 并将最终的迭代收敛值作为指标参数 θ 和联系函数 $g(\cdot)$ 的估计, 分别记为 $\hat{\theta}$ 和 $\hat{g}(\cdot)$.

下面给出指标参数 θ 估计 $\hat{\theta}$ 的渐近结果及其所需的假设条件.

(C7.3.5) Z 的密度函数 $f(z)$ 在其支撑集上恒大于零, 并且二阶偏导数有界.

(C7.3.6) $Z^{\mathrm{T}}\theta$ 的密度函数和联系函数 $g(\cdot)$ 在其支撑集上有有界连续的二阶导数.

(C7.3.7) $K_2(\cdot)$ 是定义在开区间 $(-1,1)$ 上的对称密度函数, 且导数有界.

定理 7.3.2 假设条件(C7.3.1)~条件(C7.3.7)成立, 并且当 $n\to\infty$ 时, 有 $nh_2\to 0$, $nh_2^4\to 0$ 成立, 则有

$$\sqrt{n}(\hat{\theta}-\theta)\xrightarrow{L}N\left(0,\Gamma_\theta^-\Sigma_\theta\Gamma_\theta^-\right),$$

其中 $\Gamma_\theta=E\left[\breve{Z}_\theta^{\otimes 2}g'^2(Z^{\mathrm{T}}\theta)\right]$, $\breve{Z}_\theta=Z-E[Z|Z^{\mathrm{T}}\theta]$, Γ_θ^- 为 Γ_θ 的广义逆矩阵, 并有

$$\begin{aligned}\Sigma_\theta=E\Big[&\left\{\breve{Z}_\theta g'(Z^{\mathrm{T}}\theta)-\Gamma_1\Gamma_{X|Z}^{-1}\breve{Z}_\theta\right\}(\varepsilon-U^{\mathrm{T}}\beta)\\&+\Gamma_1\Gamma_{X|Z}^{-1}\left\{U^{\otimes 2}\beta-\Sigma_{uu}\beta-U\varepsilon\right\}\Big]^{\otimes 2},\end{aligned}$$

其中 $\Gamma_1=E\left[\breve{X}_\theta\breve{Z}_\theta g'(Z^{\mathrm{T}}\theta)\right]$, $\breve{X}_\theta=X-E[X|Z^{\mathrm{T}}\theta]$. 此外, 如果 ε 和 (X,Z) 相互独立, 矩阵 Σ_θ 可简化为

$$\Sigma_\theta=\Gamma_2E(\varepsilon-U^{\mathrm{T}}\beta)^2+\Gamma_1\Gamma_{X|Z}^{-1}\left\{E\left[\left\{U^{\otimes 2}\beta-\Sigma_{uu}^{\mathrm{T}}\beta\right\}^{\otimes 2}\right]+\Sigma_{uu}\sigma^2\right\}\Gamma_{X|Z}^{-1}\Gamma_1^{\mathrm{T}},$$

其中 $\Gamma_2=E\left[\breve{Z}_\theta g'(Z^{\mathrm{T}}\theta)-\Gamma_1\Gamma_{X|Z}^{-1}\breve{X}_\theta\right]^{\otimes 2}$.

在定理 7.3.3 的渐近结果中, 如果进一步假设在给定 $Z^{\mathrm{T}}\theta$ 的情况下 X 和 Z 相互独立, 那么有 $\Gamma_1=0$ 和 $\Sigma_\theta=E\left[\breve{Z}_\theta g'(Z^{\mathrm{T}}\theta)(\varepsilon-U^{\mathrm{T}}\beta)\right]^{\otimes 2}$ 成立. 此时在大样本意义下, $\hat{\beta}_n^*$ 的估计效果并不影响 $\hat{\theta}$ 的估计. 此外, 在上述渐近结果中, Γ_1 衡量了 $\hat{\beta}_n^*$ 对估计 θ 的影响程度.

7.3.2 修正的拟似然估计方法

Liang 和 Wang (2005) 提出的两阶段估计方法主要是先将部分线性单指标模型转换为部分线性模型给出 β 的估计, 然后再转换成单指标模型给出 θ 的估计. 两阶段估计方法简单, 直观. 但是潜在的问题是如果协变量 Z 的维数过高, 那么 β 的估计会受到影响, 进而影响 θ 的估计. 为此, Liang 和 Wang (2005) 提出了修正的拟似然估计方法来克服高维问题. Liang 和 Wang (2005) 关于参数 (β,θ) 和联系函数 $g(\cdot)$ 提出极小化如下目标函数:

$$\mathrm{QL}(g,\theta,\beta,Y,Z,X)=\frac{1}{n}\sum_{i=1}^{n}\left[Y_i-W_i^{\mathrm{T}}\beta-g(Z_i^{\mathrm{T}}\theta)\right]^2-\beta^{\mathrm{T}}\Sigma_{uu}\beta. \tag{7.3.6}$$

对上述目标函数 (7.3.6) 的极小化问题, Liang 和 Wang (2005) 提出对参数 (β,θ) 和联系函数 $g(\cdot)$ 如下的迭代算法.

步骤 1 给定 (β,θ) 的一个初始的相合估计 $(\hat{\beta}_I,\hat{\theta}_{\text{ini}})$, 并正则化指标估计: $\hat{\theta}_I=\hat{\theta}_{\text{ini}}/\|\hat{\theta}_{\text{ini}}\|$.

步骤 2 记 $\hat{\Lambda}_i=Z_i^{\mathrm{T}}\hat{\theta}_I$. 关于 (a,b) 极小化下面的局部加权最小二乘目标函数

$$(\hat{a}_I(u),\hat{b}_I(u))=\arg\min_{a,b}\sum_{i=1}^{n}\left\{Y_i-W_i^{\mathrm{T}}\hat{\beta}_I-a-b(\hat{\Lambda}_i-u)\right\}^2K_{3h_3}(\hat{\Lambda}_i-u),$$

其中 $K_{3h_3}(\cdot)=h_3^{-1}K_3(\cdot/h_3)$, $K_3(\cdot)$ 为满足条件 (C7.3.7) 的核密度函数, h_3 为窗宽.

步骤 3 更新 (β,θ) 的估计, 极小化下面的全局最小二乘目标函数, 并得到更新估计 $(\hat{\beta},\hat{\theta}^*)$ 如下:

$$(\hat{\beta},\hat{\theta}^*)=\arg\min_{\beta,\theta}\left\{\frac{1}{n}\sum_{i=1}^{n}\left\{Y_i-W_i^{\mathrm{T}}\beta-\hat{a}_I(Z_i^{\mathrm{T}}\theta)\right\}^2-\beta^{\mathrm{T}}\Sigma_{uu}\beta\right\},$$

并正则化 $\hat{\theta}=\hat{\theta}^*/\|\hat{\theta}^*\|$.

步骤 4 重复迭代步骤 2 和步骤 3 直到收敛, 并把最后的估计记为 $(\hat{\beta},\hat{\theta})$.

这里需指出的是, 初始估计 $\hat{\theta}_I$ 和 $\hat{\beta}_I$ 应选取收敛速度快于 $h_3^2+\sqrt{nh_3}$ 的相合估计. 这种选取目的是为了不影响步骤 2 中的估计 $\hat{a}_I(u)$ 的渐近性质. 下面给出估计量的渐近结果.

定理 7.3.3 假设条件(C7.3.1)~条件(C7.3.7)成立, 并且当 $n\to\infty$ 时, 有 $nh_3\to 0$, $nh_3^4\to 0$ 成立, 则有

$$\sqrt{n}(\hat{\theta}-\theta)\xrightarrow{L}N\left(0,\Gamma_{\theta,L}^{-1}\Sigma_{\theta,L}\Gamma_{\theta,L}^{-1}\right),$$

其中 $\Gamma_{\theta,L}=\Gamma_\theta-\Gamma_1\Gamma_{X|\Lambda}^{-1}\Gamma_1^{\mathrm{T}}$, $\Gamma_{X|\Lambda}=E\left[\breve{X}_\theta^{\otimes 2}\right]$ 和

$$\Sigma_{\theta,L}=\Gamma_{\theta,L}E[(\varepsilon-U^{\mathrm{T}}\beta)^2]+\Gamma_1\Gamma_{X|\Lambda}^{-1}\left\{E\left[\left\{UU^{\mathrm{T}}\beta-\Sigma_{uu}^{\mathrm{T}}\beta\right\}^{\otimes 2}\right]+\Sigma_{uu}\sigma^2\right\}\Gamma_{X|\Lambda}^{-1}\Gamma_1^{\mathrm{T}}.$$

同时有

$$\sqrt{n}(\hat{\beta}-\beta)\xrightarrow{L}N\left(0,\Gamma_{\beta,L}^{-1}\Sigma_{\beta,L}\Gamma_{\beta,L}^{-1}\right),$$

其中 $\Gamma_{\beta,L}=\Gamma_{X|\Lambda}-\Gamma_1\Gamma_\theta^{-}\Gamma_1^{\mathrm{T}}$ 和

$$\Sigma_{\beta,L}=\Gamma_{\beta,L}E[(\varepsilon-U^{\mathrm{T}}\beta)^2]+E\left[\left\{UU^{\mathrm{T}}\beta-\Sigma_{uu}^{\mathrm{T}}\beta\right\}^{\otimes 2}\right]+\Sigma_{uu}\sigma^2.$$

下面考虑两阶段估计方法和拟似然估计方法关于参数 β 的估计效率问题. Liang 和 Wang (2005) 考察了模型误差 ε 和 (X,Z) 相互独立并且同方差这一特定情况下

的渐近方差比较. 定义 $\mathrm{AVAR}(\hat{\beta}_n^*)$ 为两阶段估计的渐近方差, $\mathrm{AVAR}(\hat{\beta})$ 为拟似然估计方法中的渐近方差, 那么有

$$\mathrm{AVAR}(\hat{\beta}_n^*) = E[(\varepsilon - U^{\mathrm{T}}\beta)^2]\Gamma_{X|Z}^{-1} + \Gamma_{X|Z}^{-1}\left\{E\left[\left\{UU^{\mathrm{T}}\beta - \Sigma_{uu}^{\mathrm{T}}\beta\right\}^{\otimes 2}\right] + \Sigma_{uu}\sigma^2\right\}\Gamma_{X|Z}^{-1}$$

和

$$\mathrm{AVAR}(\hat{\beta}) = E[(\varepsilon - U^{\mathrm{T}}\beta)^2]\Gamma_{\beta,L}^{-1} + \Gamma_{\beta,L}^{-1}\left\{E\left[\left\{UU^{\mathrm{T}}\beta - \Sigma_{uu}^{\mathrm{T}}\beta\right\}^{\otimes 2}\right] + \Sigma_{uu}\sigma^2\right\}\Gamma_{\beta,L}^{-1}.$$

这时, 两种估计方法所得估计量有效性的比较问题就转化为矩阵 $\Gamma_{X|Z}^{-1}$ 和矩阵 $\Gamma_{\beta,L}^{-1}$ 的比较问题. 注意到 $\Gamma_{X|Z} = E[\mathrm{Var}\{E(X|Z)\}]$, $\Gamma_{X|\Lambda} = E[\mathrm{Var}\{E(X|\Lambda)\}]$ 和

$$\mathrm{Var}(X) = \mathrm{Var}\{E(X|Z)\} + \Gamma_{X|Z} = \mathrm{Var}\{E(X|\Lambda)\} + \Gamma_{X|\Lambda},$$

$$\begin{aligned}\mathrm{Var}\{E(X|Z)\} &= \mathrm{Var}(E[E(X|Z)|\Lambda]) + E[\mathrm{Var}\{E(X|Z)|\Lambda\}] \\ &= \mathrm{Var}\{E(X|\Lambda)\} + E[\mathrm{Var}\{E(X|Z)|\Lambda\}].\end{aligned} \tag{7.3.7}$$

从式 (7.3.7) 不难看出

$$\mathrm{Var}\{E(X|Z)\} \geqslant \mathrm{Var}\{E(X|\Lambda)\}. \tag{7.3.8}$$

因此, 不等式 (7.3.8) 保证了 $\Gamma_{X|\Lambda} - \Gamma_{X|Z}$ 是半正定矩阵. 再有 $\Gamma_{\beta,L} = \Gamma_{X|\Lambda} - \Gamma_1^{\mathrm{T}}\Gamma_\theta^-\Gamma_1$ 和 $\Gamma_1^{\mathrm{T}}\Gamma_\theta^-\Gamma_1$ 是半正定矩阵, 可以得到 $\mathrm{AVAR}(\hat{\beta}_n^*) \geqslant \mathrm{AVAR}(\hat{\beta})$. 即用拟似然估计方法得到的估计 $\hat{\beta}$ 在这一特定情形下要比两阶段估计方法得到的估计 $\hat{\beta}_n^*$ 更加有效. 更详细的讨论见 Liang 和 Wang (2005).

7.3.3 经验似然方法

Zhu 和 Xue (2006) 先把指标参数 θ 转换为 $\theta = \left(\sqrt{1-\|\theta_{(1)}\|^2}, \theta_{(1)}^{\mathrm{T}}\right)^{\mathrm{T}}$, 其中 $\theta_{(1)} = (\theta_2, \cdots, \theta_q)^{\mathrm{T}}$ 是去掉第一个分量后的 $q-1$ 维向量, 然后关于参数 $(\beta, \theta_{(1)})$ 提出了纠偏的经验似然方法. Chen 和 Cui (2009) 基于 Zhu 和 Xue (2006) 提出纠偏的经验似然方法推广到了部分线性单指标 EV 模型. 为了定义经验对数似然比函数, 首先需要构造如下的辅助变量 $\xi_i(\theta_{(1)}, \beta)$, $i = 1, \cdots, n$, 即

$$\begin{aligned}\xi_i(\theta_{(1)}, \beta) = &[Y_i - W_i^{\mathrm{T}}\beta - g(Z_i^{\mathrm{T}}\theta)]\begin{pmatrix} g'(Z_i^{\mathrm{T}}\theta)J_{\theta_{(1)}}^{\mathrm{T}}(Z_i - \mu_Z(Z_i^{\mathrm{T}}\theta)) \\ W_i - \mu_W(Z_i^{\mathrm{T}}\theta)\end{pmatrix} \\ &+ \begin{pmatrix} 0 \\ \Sigma_{uu}\beta \end{pmatrix},\end{aligned} \tag{7.3.9}$$

其中 $\mu_Z(t)=E[Z|Z^{\mathrm{T}}\theta=t]$, $\mu_W(t)=E[W|Z^{\mathrm{T}}\theta=t]$ 和

$$J_{\theta_{(1)}}=\begin{pmatrix}-\dfrac{\theta_{(1)}^{\mathrm{T}}}{\sqrt{1-\|\theta_{(1)}\|^2}}\\ I_{(q-1)\times(q-1)}\end{pmatrix}.$$

此时, 辅助变量 $\xi_i\left(\theta_{(1)},\beta\right)$ 中的加项 $\left(0,\beta^{\mathrm{T}}\Sigma_{uu}\right)^{\mathrm{T}}$ 起到了纠偏的作用. 注意到当辅助变量 $\xi_i\left(\theta_{(1)},\beta\right)$ 中的 $(\theta_{(1)},\beta)$ 等于真值 $(\theta_{0(1)},\beta_0)$ 时, 有 $E\left[\xi_i\left(\theta_{0(1)},\beta_0\right)\right]=0$.

由于式 (7.3.9) 中包含未知的联系函数 $g(\cdot)$ 和导数 $g'(\cdot)$, 未知函数 $\mu_Z(t)=E[Z|Z^{\mathrm{T}}\theta=t]$ 和 $\mu_W(t)=E[W|Z^{\mathrm{T}}\theta=t]$, Chen 和 Cui (2009) 构造了如下的辅助估计量:

$$\begin{aligned}\hat{\xi}_i\left(\theta_{(1)},\beta\right)=&[Y_i-W_i^{\mathrm{T}}\beta-\hat{g}(Z_i^{\mathrm{T}}\theta;\theta,\beta)]\begin{pmatrix}\hat{g}'(Z_i^{\mathrm{T}}\theta;\theta,\beta)J_{\theta_{(1)}}^{\mathrm{T}}(Z_i-\hat{\mu}_Z(Z_i^{\mathrm{T}}\theta;\theta))\\ W_i-\hat{\mu}_W(Z_i^{\mathrm{T}}\theta;\theta)\end{pmatrix}\\&+\begin{pmatrix}0\\ \Sigma_{uu}\beta\end{pmatrix},\end{aligned}\tag{7.3.10}$$

其中 $(\hat{g}(t;\theta,\beta),\hat{g}'(t;\theta,\beta))$ 是由 Fan 和 Gijbels (1996) 提出的局部线性估计方法得到. 给定 (β,θ), 极小化局部加权最小二乘目标函数可得

$$(\hat{g}(t;\theta,\beta),\hat{g}'(t;\theta,\beta))=\arg\min_{a,b}\sum_{i=1}^n[Y_i-W_i^{\mathrm{T}}\beta-a-b(Z_i^{\mathrm{T}}\theta-t)]^2K_h\left(Z_i^{\mathrm{T}}\theta-t\right),$$

其中 $K_h(t)=K(t/h)/h$, $K(\cdot)$ 是核密度函数, h 是窗宽. 可进一步得到 $\hat{g}(t;\theta,\beta)$ 和 $\hat{g}'(t;\theta,\beta)$ 的估计分别为

$$\begin{aligned}\hat{g}(t;\theta,\beta)&=\frac{S_{20}(t;\theta,\beta)S_{01}(t;\theta,\beta)-S_{10}(t;\theta,\beta)S_{11}(t;\theta,\beta)}{S_{20}(t;\theta,\beta)S_{00}(t;\theta,\beta)-S_{10}^2(t;\theta,\beta)},\\ \hat{g}'(t;\theta,\beta)&=\frac{S_{00}(t;\theta,\beta)S_{11}(t;\theta,\beta)-S_{10}(t;\theta,\beta)S_{01}(t;\theta,\beta)}{S_{20}(t;\theta,\beta)S_{00}(t;\theta,\beta)-S_{10}^2(t;\theta,\beta)},\end{aligned}$$

这里

$$S_{l_1l_2}(t;\theta,\beta)=\sum_{i=1}^nK_h(Z_i^{\mathrm{T}}\theta-t)(Z_i^{\mathrm{T}}\theta-t)^{l_1}[Y_i-W_i^{\mathrm{T}}\beta]^{l_2},\quad l_1=0,1,2,\quad l_2=0,1.$$

类似有

$$\hat{\mu}_Z(t;\theta)=\left(\hat{\mu}_{Z_1}(t;\theta),\cdots,\hat{\mu}_{Z_q}(t;\theta)\right)^{\mathrm{T}},\quad \hat{\mu}_W(t;\theta)=\left(\hat{\mu}_{W_1}(t;\theta),\cdots,\hat{\mu}_{W_p}(t;\theta)\right)^{\mathrm{T}}$$

也是局部线性估计, 其表达式分别为

$$\hat{\mu}_{Z_s}(t;\theta)=\frac{V_{s20}(t;\theta)V_{s01}(t;\theta)-V_{s10}(t;\theta)V_{s11}(t;\theta)}{V_{s20}(t;\theta)V_{s00}(t;\theta)-V_{s10}^2(t;\theta)},$$

$$\hat{\mu}_{W_f}(t;\theta)=\frac{G_{f20}(t;\theta)G_{f01}(t;\theta)-G_{f10}(t;\theta)G_{f11}(t;\theta)}{G_{f20}(t;\theta)G_{f00}(t;\theta)-G_{f10}^2(t;\theta)},$$

这里

$$V_{sl_1l_2}(t;\theta)=\sum_{i=1}^{n}K_h(Z_i^{\mathrm{T}}\theta-t)(Z_i^{\mathrm{T}}\theta-t)^{l_1}Z_{si}^{l_2},\quad s=1,\cdots,q,$$

$$G_{fl_1l_2}(t;\theta)=\sum_{i=1}^{n}K_h(Z_i^{\mathrm{T}}\theta-t)(Z_i^{\mathrm{T}}\theta-t)^{l_1}W_{fi}^{l_2},\quad f=1,\cdots,p,l_1=0,1,2,l_2=0,1.$$

此时, 可定义 $(\theta_{(1)},\beta)$ 的经验对数似然比统计量如下:

$$\hat{\ell}_n\left(\theta_{(1)},\beta\right)=-2\max\left\{\sum_{i=1}^{n}\log(np_i)\Big|p_i\geqslant 0,\sum_{i=1}^{n}p_i=1,\sum_{i=1}^{n}p_i\hat{\xi}_i\left(\theta_{(1)},\beta\right)=0\right\}.\quad(7.3.11)$$

由 Lagrange 乘子法, 可以进一步将 $\hat{\ell}_n\left(\theta_{(1)},\beta\right)$ 表示为

$$\hat{\ell}_n\left(\theta_{(1)},\beta\right)=2\sum_{i=1}^{n}\log\left(1+\lambda^{\mathrm{T}}\hat{\xi}_i\left(\theta_{(1)},\beta\right)\right),$$

其中 λ 是 Lagrange 乘子, 由式 (7.3.12) 所决定:

$$\frac{1}{n}\sum_{i=1}^{n}\frac{\hat{\xi}_i\left(\theta_{(1)},\beta\right)}{1+\lambda^{\mathrm{T}}\hat{\xi}_i\left(\theta_{(1)},\beta\right)}=0.\quad(7.3.12)$$

为了得到经验对数似然比统计量的渐近结果, Chen 和 Cui (2009) 给出下面的假设条件.

(C7.3.8) 随机变量 $Z^{\mathrm{T}}\theta$ 的密度函数 $f_{Z^{\mathrm{T}}\theta}(t)$ 在集合 $\mathcal{T}=\{t=Z^{\mathrm{T}}\theta,z\in\mathcal{Z}\}$ 上满足一阶 Lipschiz 条件, 其中 $\mathcal{Z}$ 是随机变量 Z 的一个紧支撑集. 同时 $f_{Z^{\mathrm{T}}\theta}(t)$ 需满足

$$0<\inf_{t\in\mathcal{T}}f_{Z^{\mathrm{T}}\theta}(t)\leqslant\sup_{t\in\mathcal{T}}f_{Z^{\mathrm{T}}\theta}(t)<\infty.$$

(C7.3.9) $g(t)$, $E(Z_s|Z^{\mathrm{T}}\theta=t)$ 和 $E(W_s|Z^{\mathrm{T}}\theta=t)$ 在 $t\in\mathcal{T}$ 上有二阶连续有界的导数.

(C7.3.10) 窗宽矩阵 h 满足: 当 $n\to\infty$ 时, 有 $nh^3\to\infty$, $nh^8\to 0$ 成立.

(C7.3.11) 核函数 $K(\cdot)$ 是定义在有界支撑集上的一元连续密度函数, 且 $\int tK(t)\mathrm{d}t = 0, \int t^2K(t)\mathrm{d}t \neq 0, \int t^8K(t)\mathrm{d}t < \infty$.

(C7.3.12) 模型误差 ε 满足条件 $\sup\limits_{x,z} E[\varepsilon^2|X=x, Z=z] < \infty$, $\sup\limits_{x} E[\varepsilon^4|X=x] < \infty$; 测量误差 U 满足四阶矩条件 $E[\|U\|^4] < \infty$, 随机变量 X 的二阶矩需满足条件 $\sup\limits_{t\in\mathcal{T}} E[\|X\|^2|Z^{\mathrm{T}}\theta = t] < \infty$.

(C7.3.14) 矩阵 $E\left[(\varepsilon - \beta^{\mathrm{T}}U)(\Lambda - E[\Lambda|Z^{\mathrm{T}}\theta])^{\otimes 2}\right] + E\Gamma^{\otimes 2}$ 是正定矩阵, 其中

$$\Lambda = \begin{pmatrix} X \\ g'(Z^{\mathrm{T}}\theta)J_{\theta^{(1)}}Z \end{pmatrix}, \quad \Gamma = \begin{pmatrix} U\varepsilon - \left(U^{\otimes 2} - \Sigma_{uu}\right)\beta \\ 0 \end{pmatrix}.$$

定理 7.3.4 假设条件(C7.3.8)~条件(C7.3.14)成立, 当 $n \to \infty$ 时, 则有

$$\hat{\ell}_n\left(\theta_{0(1)}, \beta_0\right) \xrightarrow{L} \chi^2_{p+q-1},$$

其中 $(\theta_{0(1)}, \beta_0)$ 是参数 $(\theta_{(1)}, \beta)$ 的真值.

基于定理 7.3.4, 可构造 $(\theta_{(1)}, \beta)$ 的渐近置信水平为 $1-\alpha$ 的置信域, 即

$$I_{\alpha,\mathrm{EL}} = \{(\theta^*_{(1)}, \beta^*) : \hat{\ell}_n(\theta^*_{(1)}, \beta^*) \leqslant c_\alpha\},$$

其中 c_α 是 χ^2_{p+q-1} 的上 α 分位数.

7.4 部分线性单指标协变量调整模型

7.4.1 模型介绍

本节考虑部分线性单指标协变量调整模型:

$$\begin{cases} Y = X^{\mathrm{T}}\beta + g(Z^{\mathrm{T}}\theta) + \varepsilon, \\ \widetilde{Y} = \phi(U)Y, \\ \widetilde{X}_r = \psi_r(U)X_r, \quad r = 1, \cdots, p, \end{cases} \tag{7.4.1}$$

其中 Y 是不可观测的响应变量, $X = (X_1, X_2, \cdots, X_p)^{\mathrm{T}}$ 是不可观测的连续型协变量向量, $\widetilde{Y}$ 和 $\widetilde{X}_r$ 分别是观测到的响应变量和 p 维协变量, $Z = (Z_1, Z_2, \cdots, Z_q)^{\mathrm{T}}$ 是 q 维观测到的连续型协变量向量, U 是一个 1 维的观测混淆变量, $\phi(U)$ 和 $\psi_r(U)$ 是混淆变量 U 的未知失真函数, $\beta = (\beta_1, \cdots, \beta_p)^{\mathrm{T}}$ 是未知的参数向量. $g(\cdot)$ 是未知的联系函数. $\theta = (\theta_1, \cdots, \theta_q)^{\mathrm{T}}$ 是未知的指标参数向量. 不失一般性, 假设 θ 属于参数空间 $\Theta = \{\theta \in \mathbb{R}^q : \|\theta\| = 1, \theta_1 > 0\}$.

为了介绍估计程序, 需引入如下一些关于模型 (7.4.1) 基本的假设条件.

(1) $E[\phi(U)]=1, E[\psi_r(U)]=1, r=1,\cdots,p$.

(2) (X,Z,U,ε) 是相互独立的, 且 (Y,U) 也是相互独立的.

(3) $E(\varepsilon)=0, \mathrm{Var}(\varepsilon)=\sigma^2<\infty$.

7.4.2 模型转换和估计思想

把 $Y=\widetilde{Y}/\phi(U), X_r=\widetilde{X}_r/\psi_r(U), r=1,\cdots,p$, 代入到模型 (7.4.1) 中的第一式, 把模型 (7.4.1) 转换为下面的变系数单指标模型:

$$\widetilde{Y}=\widetilde{X}^{\mathrm{T}}\beta(U)+\phi(U)g(Z^{\mathrm{T}}\theta)+\phi(U)\varepsilon, \tag{7.4.2}$$

这里 $\beta(U)=\big(\beta_1\phi(U)/\psi_1(U),\beta_2\phi(U)/\psi_2(U),\cdots,\beta_p\phi(U)/\psi_p(U)\big)^{\mathrm{T}}$. 值得指出的是, 这里的转换模型方法类似于第 2 章介绍的 Sentürk 和 Müller (2009) 提出的转换方法. Zhang 等 (2013b) 把 Xia 等 (2002) 提出的 MAVE 方法推广到部分线性单指标协变量调整模型 (7.4.1) 的研究, 给出了模型 (7.4.2) 中未知变系数函数 $\beta(\cdot)$ 和指标参数 θ 的估计. 进而得到模型 (7.4.1) 中未知参数 (β,θ) 和未知联系函数 $g(\cdot)$ 的估计.

假设 $\Big\{(\widetilde{X}_i,Z_i,U_i,\widetilde{Y}_i), i=1,\cdots,n\Big\}$是一组 i.i.d. 的观察样本, 来自于模型 (7.4.1). 关于未知函数 $\phi(u)g(z^{\mathrm{T}}\theta)$ 的估计, 采用局部线性逼近方法. 当样本值 $(Z_i^{\mathrm{T}}\theta,U_i)$ 很接近 $(Z_j^{\mathrm{T}}\theta,U_j)$ 时, 近似地有

$$\begin{aligned}\beta_s(U_i)&\approx\beta_s(U_j)+\beta_s'(U_j)(U_i-U_j)\\&=:b_{j,s}+c_{j,s}(U_i-U_j),\quad s=1,\cdots,p\end{aligned}$$

和

$$\begin{aligned}\phi(U_i)g(Z_i^{\mathrm{T}}\theta)&\approx\phi(U_j)g(Z_j^{\mathrm{T}}\theta)+\phi(U_j)g'(Z_j^{\mathrm{T}}\theta)\theta^{\mathrm{T}}(Z_i-Z_j)+\phi'(U_j)g(Z_j^{\mathrm{T}}\theta)(U_i-U_j)\\&=:a_j+d_{1j}\theta^{\mathrm{T}}(Z_i-Z_j)+d_{2j}(U_i-U_j).\end{aligned}$$

记 $b_j=(b_{j,1},\cdots,b_{j,p})^{\mathrm{T}}$, Zhang 等 (2013b) 构造了如下的目标函数:

$$\min_{\theta^{\mathrm{T}}\theta=1}\sum_{j=1}^{n}\sum_{i=1}^{n}\Big\{\widetilde{Y}_i-\widetilde{X}_i^{\mathrm{T}}b_j-a_j-d_{1j}\theta^{\mathrm{T}}(Z_i-Z_j)-d_{2j}(U_i-U_j)\Big\}^2\omega_i(Z_j,U_j), \tag{7.4.3}$$

其中 $\omega_i(Z_j,U_j)$ 是权函数. 如何选择该权函数, 将在 7.4.3 节详细讨论. 当 θ 已知时, 目标函数 (7.4.3) 是关于 (a_j,b_j,d_{1j},d_{2j}) 的局部线性估计问题; 当 $(a_j,b_j,\ d_{1j},d_{2j})$ 已知时, 目标函数 (7.4.3) 是关于参数 θ 的加权最小二乘估计问题. 因此, 对目标函数 (7.4.3) 极小化可采用迭代的方法. 具体而言, 在未知项 (a_j,b_j,d_{1j},d_{2j}) 与未知指标参数 θ 之间进行迭代. 值得一提的是, 本节提出的迭代估计方法与 Xia 等 (2002)

提出的 MAVE 方法很相似. 不难看出, 对目标函数 (7.4.3) 极小化求解问题, 可归结为一个二次规划问题, 并在每一个迭代步都有估计量的显示表达式. 下面给出详细的迭代估计步骤.

7.4.3 指标参数 θ 的估计

1. *指标参数 θ 的初始估计*

基于 U 和 Z 相互独立这个假设条件, 权函数 $\omega_i(Z_j,U_j)$ 可取为乘积核函数. 定义为

$$\omega_i(Z_j,U_j)=\frac{I_n(Z_j,U_j)H_{h_1}(Z_i-Z_j)L_{h_2}(U_i-U_j)}{\sum\limits_{i=1}^{n}H_{h_1}(Z_i-Z_j)L_{h_2}(U_i-U_j)},$$

其中 $H(\cdot)$ 是一个 q 维变量的多元密度函数, $H_{h_1}(\cdot)=h_1^{-q}H(\cdot/h_1)$, $L(\cdot)$ 是一个单变量密度函数, $L_{h_2}(\cdot)=h_2^{-1}L(\cdot/h_2)$, h_1 和 h_2 是两个不同的窗宽. 示性函数 $I_n(Z_j,U_j)$ 选取为 $I\left\{n^{-1}\sum\limits_{i=1}^{n}H_{h_1}(Z_i-Z_j)L_{h_2}(U_i-U_j)>c_0c_1\right\}$, 其中$c_0$ 和 c_1 是两个预先给定的常数. 在这里需指出, 虽然使用多元密度函数 $H(\cdot)$ 会遭遇 “维数祸根” 的问题, 但是这里只是为了给出 θ 的初始估计.

为了简单, 定义 $\mu_j=d_{1j}\theta$, $\mathcal{M}_{ij}=\left(1,\widetilde{X}_i^{\mathrm{T}},(U_i-U_j),(Z_i-Z_j)^{\mathrm{T}}\right)^{\mathrm{T}}$ 和 $\omega_{ij}=\omega_i(Z_j,U_j)$. 关于 (a_j,b_j,d_{2j},μ_j) 极小化目标函数 (7.4.3), 可以得到

$$\left(\hat{a}_j,\hat{b}_j,\hat{d}_{2j},\hat{\mu}_j\right)^{\mathrm{T}}=\left\{\sum_{i=1}^{n}\mathcal{M}_{ij}\mathcal{M}_{ij}^{\mathrm{T}}\omega_{ij}\right\}^{-1}\sum_{i=1}^{n}\mathcal{M}_{ij}\widetilde{Y}_i\omega_{ij}.$$

记 $\hat{\mathcal{V}}_n=(\hat{\mu}_1,\hat{\mu}_2,\cdots,\hat{\mu}_n)^{\mathrm{T}}$, 再定义 $\hat{\varsigma}$ 是矩阵 $\hat{\mathcal{V}}_n\hat{\mathcal{V}}_n^{\mathrm{T}}/n$ 最大特征值所对应的特征向量: 即

$$\left(\frac{1}{n}\hat{\mathcal{V}}_n\hat{\mathcal{V}}_n^{\mathrm{T}}\right)\hat{\varsigma}=\lambda_{\max}\hat{\varsigma},$$

其中 $\lambda_{\max}$ 是最大的特征值, $\hat{\varsigma}$ 是最大特征值对应的特征向量.

最后给出指标参数 θ 的初始估计. 注意到 θ 的第一个分量为正数, 那么 θ 的初始估计为

$$\hat{\theta}_{\mathrm{ini}}=\mathrm{sgn}(\hat{\varsigma}_1)\hat{\varsigma},$$

其中, $\hat{\varsigma}_1$ 是 $\hat{\varsigma}$ 的第一个分量.

这里之所以取最大特征值所对应的特征向量 $\hat{\varsigma}$ 作为构造指标参数 θ 初始估计是因为: 当 n 趋于无穷时, $\hat{\mu}_j-\phi(U_j)g'(Z_j^{\mathrm{T}}\theta)\theta\xrightarrow{P}0$. 也就是说, $\hat{\mu}_j$ 依概率收敛到 θ 的倍数. 这样, $\hat{\varsigma}$ 依概率收敛到 θ.

2. 指标参数 θ 的精细估计

为了使 θ 的最终估计精度提高, 在给出初始估计 $\hat{\theta}_{\rm ini}$ 之后, 下面考虑 θ 的精细估计, 算法如下.

步骤 1　给定初始估计 θ^*, 定义 $\mathcal{M}_{ij,\theta^*} = \left(1, \widetilde{X}_i^{\rm T}, (Z_i - Z_j)^{\rm T}\theta^*, (U_i - U_j)\right)^{\rm T}$ 和

$$\omega_{ij}^{\theta^*} = \frac{I_n^{\theta^*}(Z_j, U_j) K_h(Z_i^{\rm T}\theta^* - Z_j^{\rm T}\theta^*) K_h(U_i - U_j)}{\sum_{i=1}^{n} K_h(Z_i^{\rm T}\theta^* - Z_j^{\rm T}\theta^*) K_h(U_i - U_j)},$$

其中 $K(\cdot)$ 是一个单变量密度函数, h 是窗宽, 示性函数定义为

$$I_n^{\theta^*}(Z_j, U_j) = I\left\{n^{-1}\sum_{i=1}^{n} K_h(Z_i^{\rm T}\theta^* - Z_j^{\rm T}\theta^*) K_h(U_i - U_j) > c_0 c_1\right\},$$

则有

$$\left(\hat{a}_{j,\theta^*}, \hat{b}_{j,\theta^*}, \hat{d}_{1j,\theta^*}, \hat{d}_{2j,\theta^*}\right)^{\rm T} = \left\{\sum_{i=1}^{n} \mathcal{M}_{ij,\theta^*} \mathcal{M}_{ij,\theta^*}^{\rm T} \omega_{ij}^{\theta^*}\right\}^{-1} \sum_{i=1}^{n} \mathcal{M}_{ij,\theta^*} \widetilde{Y}_i \omega_{ij}^{\theta^*}. \quad (7.4.4)$$

步骤 2　由式 (7.4.4) 得到 $(\hat{a}_{j,\theta^*}, \hat{b}_{j,\theta^*}, \hat{d}_{1j,\theta^*}, \hat{d}_{2j,\theta^*}, \theta^*)$, 并记

$$\widetilde{Y}_{ij}^* = \widetilde{Y}_i - \hat{a}_{j,\theta^*} - \widetilde{X}_i^{\rm T}\hat{b}_{1j,\theta^*} - \hat{d}_{2j,\theta^*}(U_i - U_j),$$

可的 θ 的更新估计为

$$\hat{\theta} = \left\{\sum_{j=1}^{n}\sum_{i=1}^{n} \hat{d}_{1j,\theta^*}^2 (Z_i - Z_j)(Z_i - Z_j)^{\rm T} \omega_{ij}^{\theta^*}\right\}^{-1} \sum_{j=1}^{n}\sum_{i=1}^{n} \hat{d}_{1j,\theta^*}(Z_i - Z_j)\widetilde{Y}_{ij}^* \omega_{ij}^{\theta^*}. \quad (7.4.5)$$

对得到的估计 $\hat{\theta}$ 进行正则化, 则有 $\hat{\theta} = {\rm sgn}(\hat{\theta}_1)\hat{\theta}/\|\hat{\theta}\|$, 这里 $\hat{\theta}_1$ 表示 $\hat{\theta}$ 的第一个分量.

步骤 3　在式 (7.4.4) 和式 (7.4.5) 之间迭代, 直到收敛. 即, 由步骤 1 和步骤 2 进行迭代, 最后得到的收敛估计就是指标参数 θ 的最终估计量, 记为 $\hat{\theta}$.

在上述算法中可先用初始估计 $\hat{\theta}_{\rm ini}$ 作为步骤 1 中的 θ^*, 接着由式 (7.4.4) 得到 $(\hat{a}_{j,\theta^*}, \hat{b}_{j,\theta^*}, \hat{d}_{1j,\theta^*}, \hat{d}_{2j,\theta^*})$, 并由式 (7.4.5) 来更新 θ 的估计.

为了给出指标参数 θ 的最终估计 $\hat{\theta}$ 的渐近结果, 列出下面的假设条件.

(C7.4.1) 依概率 1, 变量 Z 在其支撑集 $\mathcal{Z}$ 上取值, 其密度函数 $f_Z(z)$ 的三阶导数有界, 并且存在正常数 c_0, 使得区域 $\{z : f_Z(z) > c_0\}$ 不是空集.

(C7.4.2) 对满足可识别条件 $\|\theta\| = 1$ 的所有指标参数 θ, $Z^{\rm T}\theta$ 的密度函数 $f_{Z^{\rm T}\theta}(t)$ 的三阶导数有界, 并且存在正常数 c_0, 使得 $\{z : f_{Z^{\rm T}\theta}(z^{\rm T}\theta) > c_0\}$ 不是空集.

(C7.4.3) 存在一个正常数 c_1, 使得当混淆变量 U 在其取值的支撑集 $\mathcal{U}$ 上, U 的密度函数 $f_U(u)$ 满足 $f_U(u) > c_1$.

(C7.4.4) 联系函数 $g(\cdot)$ 的三阶导数连续有界, 并且在其支撑集上不恒为常数.

(C7.4.5) 模型误差满足 $E\varepsilon^4 < \infty$, 并且 $(X^{\mathrm{T}}, Z^{\mathrm{T}})^{\mathrm{T}}$ 的协方差矩阵正定. 此外, 协变量 (X, Z) 需满足下面的矩条件: $E(X_r) \neq 0$, $EX_r^4 < \infty$, $r = 1, \cdots, p$, $EZ_l^4 < \infty$, $l = 1, \cdots, q$.

(C7.4.6) 核函数 $H(\cdot)$ 是一个定义在有界支撑上, 连续且关于原点对称的多元密度函数, 同时要求 $H(\cdot)$ 的一阶偏导数有界, 并有

$$\int z_r^2 H(z_1, \cdots, z_q)\mathrm{d}z_r \neq 0, \quad \int |z_r|^j H(z_1, \cdots, z_q)\mathrm{d}z_r < \infty, \quad j=1,2,3, r = 1, \cdots, q.$$

核函数 $L(\cdot)$ 和 $K(\cdot)$ 是定义在有界支撑集上的一元连续且关于原点对称的密度函数, 满足

$$\int t^2 L(t)\mathrm{d}t \neq 0, \quad \int t^2 K(t)\mathrm{d}t \neq 0, \quad \int |t|^j L(t)\mathrm{d}t < \infty, \quad \int |t|^j K(t)\mathrm{d}t < \infty, \quad j=1,2,3,$$

并且 $L(\cdot)$ 和 $K(\cdot)$ 的一阶导数有界. 此外, 核函数 $K(\cdot)$ 的二阶导数在其定义域内有界.

(C7.4.7) 失真函数 $\phi(u)$, $\psi_r(u)$ 的三阶导数连续有界, 并且当 $u \in \mathcal{U}$ 时有, 函数 $\phi(u)$, $\psi_r(u)$ 都不为 0, $r = 1, \cdots, p$. 此外, $E\phi^4(U) < \infty$, $E\psi_r^2(U) < \infty$, $r = 1, \cdots, p$.

(C7.4.8) 函数 $\xi_\theta(t)$, $\varsigma_\theta(t)$, $S_\theta(t)$,$\mathcal{T}_\theta(t)$, $\mathcal{W}_\theta(t)$, $\mathcal{V}_\theta(t)$ 关于 t 存在三阶连续的导数, 并且三阶导数在 $\mathcal{C}$ 上满足一致 Lipschiz 连续性条件, 其中 $\mathcal{C} = \{t = Z^{\mathrm{T}}\theta : z \in \mathcal{Z} \subset \mathbb{R}^q, \theta \in \Theta \subset \mathbb{R}^q\}$, $\mathcal{Z}$ 是 Z 的一个紧支撑集. 此外, $\mathcal{W}_\theta(t)$ 在 $\mathcal{C}$ 上可逆.

条件 (C7.4.1) 是关于密度函数 $f_Z(z)$, $f_{Z^{\mathrm{T}}\theta}(z^{\mathrm{T}}\theta)$ 和 $f_U(u)$ 的技术条件, 主要是为了使用核光滑方法估计这些密度函数时避免分母为 0 的情况. 条件 (C7.4.2) 是关于联系函数 $g(\cdot)$ 的光滑性假设, 见 Liang 等 (2010) 和 Xia 和 Härdle (2006). 条件 (C7.4.3)~(C7.4.5) 是定理 7.4.1~定理 7.4.3 中用于保证渐近分布有意义的假设. 条件 (C7.4.6) 是常用的非参数光滑估计的常用假设. 条件 (C7.4.7) 是关于失真函数的光滑性条件假设. 该条件也是目前文献中关于失真函数的常用假设, 见 Sentürk 和 Müller (2005a, 2005b, 2006, 2009), Cui 等 (2009). $E\psi_r^2(U) < \infty$ 是为了保证 $E\{\beta(U)\beta^{\mathrm{T}}(U)\} < \infty$, $E\phi^4(U)$ 有限是为了保证定理 7.4.3 渐近方差有限. 条件 (C7.4.8) 是非参数估计中需要假定的光滑性条件, 其目的为了保证 (β, θ) 的估计有 $\sqrt{n}$ 的相合估计.

定理 7.4.1 假设条件(C7.4.1)~条件(C7.4.3)和条件(C7.4.5)成立, 并且当 $n \to \infty$ 时, $h_1 \to 0$, $h_2 \to 0$, $h_2^3/h_1 \to 0$ 和 $nh_1^{q+2}h_2 \to \infty$ 成立, 则有 $\hat{\theta}_{\mathrm{ini}} \xrightarrow{P} \theta$.

定理 7.4.1 表明, 即使用多元核函数 $H(\cdot)$, 仍可以得到 θ 的一个相合的初始估计 $\hat{\theta}_{\rm ini}$. 正如 Xia 和 Härdle (2006) 使用的 MAVE 算法一样, 使用该初始估计并结合步骤 1 和步骤 2 的迭代算法, 可最终得到 θ 的 $\sqrt{n}$ 相合估计.

定义 $\xi_\theta(t)=E(X|Z^{\rm T}\theta=t)$, $\varsigma_\theta(t)=E\big(Z|Z^{\rm T}\theta=t\big)$, $S_\theta(t)=E\big(X^{\otimes 2}|Z^{\rm T}\theta=t\big)$,$\mathcal{T}_\theta(t)=E\big(ZX^{\rm T}|Z^{\rm T}\theta=t\big)$, $\mathcal{W}_\theta(t)=S_\theta(t)-\xi_\theta^{\otimes 2}(t)$, $\mathcal{V}_\theta(t)=E\big(Z^{\otimes 2}|Z^{\rm T}\theta=t\big)-\varsigma_\theta^{\otimes 2}(t)$, $\mathcal{D}_\theta=E\Big[\phi(U)^2\big\{g'(Z^{\rm T}\theta)\big\}^2\mathcal{V}_\theta(Z^{\rm T}\theta)\mathcal{I}^\theta(Z,U)\Big]$ 和 $\mathcal{I}_\theta(z,u)=I\{(z,u)\in\mathcal{Z}\times\mathcal{U}: f_{Z^{\rm T}\theta}(z^{\rm T}\theta)f_u(U)>c_0c_1\}$.

定理 7.4.2 假设条件(C7.4.1)~条件(C7.4.8)成立, 当 $n\to\infty$ 时, $nh^2\to\infty$, $nh^6\to 0$ 成立, 则有 $\sqrt{n}\big(\hat{\theta}-\theta\big)$ 依分布收敛到均值为 0, 协方差阵为 Σ_θ 的正态分布, 其中

$$\Sigma_\theta=\frac{\sigma^2}{4}\mathcal{D}_\theta^-E\Big[\Big\{\big(\mathcal{T}_\theta(Z^{\rm T}\theta)-\varsigma_\theta(Z^{\rm T}\theta)\xi_\theta^{\rm T}(Z^{\rm T}\theta)\big)\mathcal{W}_\theta^{-1}(Z^{\rm T}\theta)\big(X-\xi_\theta(Z^{\rm T}\theta)\big)$$
$$+\big(\varsigma_\theta(Z^{\rm T}\theta)-Z\big)\Big\}\phi^2(U)g'(Z^{\rm T}\theta)\mathcal{I}_\theta(Z,U)\Big]^{\otimes 2}\mathcal{D}_\theta^-,$$

这里 $\mathcal{D}_\theta^-$ 表示矩阵 $\mathcal{D}_\theta$ 的广义逆.

由于 $\mathcal{V}_{\theta_0}(t)$ 是在给定 Z 自身的线性约束下 Z 的条件协方差矩阵, 进而可以得到如下等式 $\theta^{\rm T}\mathcal{V}_\theta(Z^{\rm T}\theta)\theta=0$ 成立. 所以, 矩阵 $\mathcal{D}_\theta$ 与矩阵 Σ_θ 的秩至多是 $q-1$ 而不是 q. 因此, 定理 7.4.2 中出现了矩阵 $\mathcal{D}_\theta$ 的广义逆 $\mathcal{D}_\theta^-$. 类似的结果可见 Härdle 等 (1993), Xia 和 Härdle (2006).

7.4.4 参数 β 的估计

注意到 $E\widetilde{Y}=EY$, $E\widetilde{X}=EX$, 以及 U 和 (X,Y) 相互独立, 则在总体情形下有: $E\big\{\beta_r(U)\widetilde{X}_r\big\}=\beta_rEX_r=\beta_rE\widetilde{X}_r$, $r=1,\cdots,p$. Sentürk 和 Müller (2005a) 基于假设条件 (C7.4.5), 可得

$$\beta_r=E\beta_r(U)\widetilde{X}_r/E\widetilde{X}_r. \tag{7.4.6}$$

不同于 Sentürk 和 Müller (2005a, 2006) 提出的 binning 方法, Zhang 等 (2013b) 基于式 (7.4.6) 提出如下 β_r 的估计:

$$\hat{\beta}_r=\sum_{j=1}^n\widetilde{X}_{j,r}\hat{b}_{j,\hat{\theta},r}\Big/\sum_{j=1}^n\widetilde{X}_{j,r}, \tag{7.4.7}$$

其中 $\hat{b}_{j,\hat{\theta},r}$ 是 $\hat{b}_{j,\hat{\theta}}$ 的第 r 个分量, $\widetilde{X}_{j,r}$ 是 $\widetilde{X}_j$ 的第 r 个分量, $j=1,\cdots,n$, $r=1,\cdots,p$. 此处 $\hat{b}_{j,\hat{\theta},r}$ 是 7.4.3 节中基于步骤 1 和步骤 2 迭代得到的收敛估计.

记 $F_\theta(Z^{\rm T}\theta)={\rm diag}\left(\dfrac{\xi_{\theta,1}(Z^{\rm T}\theta)}{EX_1},\cdots,\dfrac{\xi_{\theta,p}(Z^{\rm T}\theta)}{EX_p}\right)$, $L_\beta={\rm diag}\left(\dfrac{\beta_1}{EX_1},\cdots,\dfrac{\beta_p}{EX_p}\right)$,

其中 $\xi_{\theta,j}(t)$ 是 $\xi_\theta(t)$ 的第 j 个分量, $j=1,\cdots,p$.

定理 7.4.3 假设条件(C7.4.1)~条件(C7.4.8)成立, 并且当 $n\to\infty$ 时, $nh^2\to\infty$, $nh^6\to 0$ 成立, 则有 $\sqrt{n}(\hat\beta-\beta)$ 依分布收敛到均值为 0, 协方差阵为 Σ_β 的正态分布, 其中

$$\Sigma_\beta = L_\beta E\Big\{\big(\phi(U)I_p-\psi(U)\big)\big\{EX^{\otimes 2}\big\}\big(\phi(U)I_p-\psi(U)\big)\Big\}L_\beta$$
$$+\sigma^2 E\Big\{\phi^2(U)F_\theta(Z^{\mathrm T}\theta)\mathcal{W}_\theta^{-1}(Z^{\mathrm T}\theta)F_\theta(Z^{\mathrm T}\theta)\mathcal{I}_\theta(Z_j,U_j)\Big\}.$$

7.4.5 变量选择

1. 变量选择程序和渐近性质

注意到: $\beta(U)=\mathrm{diag}\left(\dfrac{\phi(U)}{\psi_1(U)},\dfrac{\phi(U)}{\psi_2(U)},\cdots,\dfrac{\phi(U)}{\psi_p(U)}\right)\beta$, 即 $\beta_r(u)=\beta_r\phi(u)/\psi_r(u)$, $r=1,\cdots,p$. 因此当条件 (C7.4.6) 成立时, 在 u 属于 U 的支撑集内有 $\phi(u)/\psi_r(u)$ 的取值恒不为 0; 这意味着 $\beta_r(u)=0$ 当且仅当 $\beta_r=0$. 因此, 挑选 β 中的零元素等价于挑选 $\beta(u)$ 中的零元素. 这时, 变量选择这一问题可通过如何得到矩阵 $E\{\beta(U)\beta(U)^{\mathrm T}\}$ 的第一稀疏主成分 (the first sparse principal component, SPC) 来解决.

构造 $E\{\beta(U)\beta(U)^{\mathrm T}\}$ 的估计如下:

$$\hat E\Big\{\beta(U)\beta(U)^{\mathrm T}\Big\}=\frac{1}{n}\sum_{j=1}^n \hat b_{j,\hat\theta}\hat b_{j,\hat\theta}^{\mathrm T},$$

其中$\{\hat b_{1,\hat\theta},\cdots,\hat b_{n,\hat\theta}\}$ 是 7.4.3 节中的步骤 1 和步骤 2 迭代得到的收敛估计.

为了得到第一稀疏主成分的估计, 首先介绍矩阵 $\hat E\Big\{\beta(U)\beta(U)^{\mathrm T}\Big\}$ 第一主成分的估计 $\hat\eta_n$. 该估计可通过如下的特征值分解来得到:

$$\left\{\frac{1}{n}\sum_{j=1}^n \hat b_{j,\hat\theta}\hat b_{j,\hat\theta}^{\mathrm T}\right\}\hat\eta_n=\hat\lambda\hat\eta_n, \tag{7.4.8}$$

其中 $\hat\lambda$ 是矩阵 $\dfrac{1}{n}\sum\limits_{j=1}^n \hat b_{j,\hat\theta}\hat b_{j,\hat\theta}^{\mathrm T}$ 的最大特征值. 注意到上述特征值分解 (7.4.8) 可通过极小化下面的最小二次平方函数来实现 (Li, 2007):

$$\hat\eta_n=\arg\min_v\sum_{i=1}^p\|m_i-vv^{\mathrm T}m_i\|^2,\quad 约束条件是\ \ v^{\mathrm T}v=1, \tag{7.4.9}$$

其中 v 是一个 $p\times 1$ 的向量, m_i 是矩阵 $\left\{\dfrac{1}{n}\sum\limits_{j=1}^n \hat b_{j,\hat\theta}\hat b_{j,\hat\theta}^{\mathrm T}\right\}^{1/2}$ 的第 i 列, $i=1,\cdots,p$.

这里矩阵 $\left\{\frac{1}{n}\sum_{j=1}^{n}\hat{b}_{j,\hat{\theta}}\hat{b}_{j,\hat{\theta}}^{\mathrm{T}}\right\}^{1/2}$ 满足

$$\left\{\frac{1}{n}\sum_{j=1}^{n}\hat{b}_{j,\hat{\theta}}\hat{b}_{j,\hat{\theta}}^{\mathrm{T}}\right\}^{1/2}\left\{\frac{1}{n}\sum_{j=1}^{n}\hat{b}_{j,\hat{\theta}}\hat{b}_{j,\hat{\theta}}^{\mathrm{T}}\right\}^{1/2}=\frac{1}{n}\sum_{j=1}^{n}\hat{b}_{j,\hat{\theta}}\hat{b}_{j,\hat{\theta}}^{\mathrm{T}}.$$

为了得到第一稀疏主成分, 基于极小化目标函数 (7.4.9), Chen 等 (2010) 提出的坐标独立的稀疏充分降维 (coordinate-independent sparse sufficient dimension reduction, CISE) 方法, 即关于 v 极小化如下目标函数

$$\breve{\eta}_n=\arg\min_{v}\left\{\sum_{i=1}^{p}\|m_i-vv^{\mathrm{T}}m_i\|^2+\sum_{r=1}^{p}\alpha_r|v_r|\right\},\tag{7.4.10}$$

其中 $\alpha_r\geqslant 0$ 是惩罚参数.

在目标函数 (7.4.10) 中采用了绝对值函数 $|\cdot|$ 来得到稀疏的解, 从而达到变量选择的目的. Chen 等 (2010) 也指出, 可以选择其他在零点不可导的正凸函数来达到变量选择的目的. 其实, 零点不可导的这个特性恰好可以使 $\breve{\eta}_n$ 中的一些数值上很小的分量被压缩为零, 也就实现了变量选择的目的. 值得注意的是, 这里对不同的分量采用了不同的惩罚参数, 这样做的目的是为了尽可能选出正确的变量, 从而减小模型偏差. 最后由式 (7.4.10) 得到的解 $\breve{\eta}_n$ 便是矩阵 $E\{\beta(U)\beta(U)^{\mathrm{T}}\}$ 的第一稀疏主成分的估计, 进而可定义线性部分 β 的稀疏估计如下:

$$\hat{\beta}_r^s=\sum_{j=1}^{n}\widetilde{X}_{j,r}\hat{b}_{j,\hat{\theta},r}I\{\breve{\eta}_{n,r}\neq 0\}\Big/\sum_{j=1}^{n}\widetilde{X}_{j,r},\quad r=1,\cdots,p,\tag{7.4.11}$$

其中 $I\{\cdot\}$ 表示示性函数, $\breve{\eta}_{n,r}$ 是 $\breve{\eta}_n$ 的第 r 个分量.

为了完成上面的估计, 下面介绍详细的算法.

极小化目标函数 (7.4.10) 与极小化下面的目标函数等价

$$\mathcal{L}=-v^{\mathrm{T}}\left\{\frac{1}{n}\sum_{i=1}^{n}\hat{b}_{j,\hat{\theta}}\hat{b}_{j,\hat{\theta}}^{\mathrm{T}}\right\}v+\sum_{r=1}^{p}\alpha_r|v_r|.\tag{7.4.12}$$

通过极小化目标函数 (7.4.12) 进行求解, 可采用 Chen 等 (2010) 提出的算法.

首先使用由 Fan 和 Li (2001) 提出的局部二次逼近的思路来处理绝对值函数 $|\cdot|$, 从而在算法上克服绝对值函数在零点不可导的问题. 记 $\rho(v)=\sum_{r=1}^{p}\alpha_r|v_r|$. 假设 v 的每个分量都不为零, 那么 $\rho(v)$ 关于 v 的一阶导数为 $\frac{\partial\rho(v)}{\partial v}=\mathrm{diag}\left(\frac{\alpha_1}{|v_1|},\cdots,\frac{\alpha_p}{|v_p|}\right)v$.

由局部二次逼近方法, $\rho(v)$ 的一阶导数在 $v^{(0)}$ 处的取值可近似表示为

$$\frac{\partial\rho(v)}{\partial v}\approx \mathrm{diag}\left(\frac{\alpha_1}{\left|v_1^{(0)}\right|},\cdots,\frac{\alpha_p}{\left|v_p^{(0)}\right|}\right)v=\mathrm{N}^{(0)}v. \tag{7.4.13}$$

从而, 利用二阶 Taylor 展式, 在 $v^{(0)}$ 附近可以得到

$$\rho(v)\approx\frac{1}{2}v^{\mathrm{T}}\mathrm{N}^{(0)}v+c_0^*, \tag{7.4.14}$$

其中 c_0^* 为某个常数. 有了上面的准备工作, 下面可详细介绍迭代算法, 并如何求解估计 $\breve{\eta}_n$.

步骤 1 给定一个初值向量 $v^{(0)}$, 用于得到式 (7.4.14) 的近似表达值. 关于 v 来极小化下面目标函数, 得到一个更新值 $v^{(1)}$,

$$\mathcal{L}^{(1)}=-v^{\mathrm{T}}\left\{\frac{1}{n}\sum_{i=1}^{n}\hat{b}_{j,\hat{\theta}}\hat{b}_{j,\hat{\theta}}^{\mathrm{T}}\right\}v+\frac{1}{2}v^{\mathrm{T}}\mathrm{N}^{(0)}v. \tag{7.4.15}$$

$v^{(1)}$ 实际上是矩阵 $\frac{1}{n}\sum\limits_{i=1}^{n}\hat{b}_{j,\hat{\theta}}\hat{b}_{j,\hat{\theta}}^{\mathrm{T}}-\frac{1}{2}\mathrm{N}^{(0)}$ 的最大特征值所对应的特征向量.

步骤 2 将步骤 1 中得到的 $v^{(1)}$ 作为初始值, 由式 (7.4.14) 来更新 $\rho(v)$ 的近似值, 进而迭代式 (7.4.13)~式 (7.4.15), 直到收敛. 在迭代过程中, 可预先设定一个非常小的正数, 如 $\delta=10^{-6}$. 如果第 k 次迭代过程中 $\left|v_r^{(k)}\right|<\delta$, 那么就将 v 的第 r 个分量去掉. 关于初始向量 $v^{(0)}$ 的选取, 可直接使用由式 (7.4.9) 中得到的特征向量 $\hat{\eta}_n$.

定理 7.4.4 定义集合 $\mathcal{A}=\{r:\beta_r\neq 0\}$ 和集合 $\mathcal{A}_n=\{r:\breve{\eta}_{n,r}\neq 0\}$. 在定理 7.4.2 的条件下, 有

(1) $\frac{1}{n}\sum\limits_{j=1}^{n}\hat{b}_{j,\hat{\theta}}\hat{b}_{j,\hat{\theta}}^{\mathrm{T}}-E\{\beta(U)\beta(U)^{\mathrm{T}}\}=O_P(n^{-1/2})$.

(2) 取 $\alpha_r=\alpha_0|\hat{\eta}_{n,r}|^{-\varpi}$, 其中 ϖ 满足 $\varpi>0$, α_0 满足 $\alpha_0\to 0$ 和 $\alpha_0 n^{\varpi/2}\to\infty$, 则有 $P(\mathcal{A}_n=\mathcal{A})\to 1$.

定理 7.4.4 (1) 说明估计 $\frac{1}{n}\sum\limits_{j=1}^{n}\hat{b}_{j,\hat{\theta}}\hat{b}_{j,\hat{\theta}}^{\mathrm{T}}$ 是 $E\{\beta(U)\beta^{\mathrm{T}}(U)\}$ 的一个 $\sqrt{n}$ 相合估计. 由 Kato (1983) 的扰动理论可知, 特征向量 $\hat{\eta}_n$ 也是 $\sqrt{n}$ 相合的. 因此可选用 $\hat{\eta}_n$ 作为初始值 $v^{(0)}$.

如果假设第 r 个分量 $\beta_r=0$, 那么矩阵 $E\left\{\beta(U)\beta(U)^{\mathrm{T}}\right\}$ 的第 r 行和第 r 列都是零元素, 这就意味着第一主成分 η 的第 r 分量 η_r 也为 0. 因此根据定理 7.4.4(1) 中的结论, 可得 $\hat{\eta}_{n,r}=\eta_r+O_P(n^{-1/2})=O_P(n^{-1/2})$, 再由定理条件 $\alpha_0 n^{\varpi/2}\to\infty$, 有 $\alpha_r\xrightarrow{P}\infty$, 进而把 $\breve{\eta}_{n,r}$ 压缩成零. 另外, 如果 $\beta_{r'}$ 非零并且 $\eta_{r'}$ 也非零, 那么有

$\hat{\eta}_{n,r'} = \eta_{r'} + O_P(n^{-1/2}) = O_P(1)$. 再由定理条件 $\alpha_0 \to 0$, 可以得到 $\alpha_{r'} \xrightarrow{P} 0$. 从而在渐近意义上, $\alpha_{r'}$ 对 $\breve{\eta}_{n,r'}$ 不进行惩罚. 因此选择 $\hat{\eta}_n$ 作为初始值 $v^{(0)}$ 能够自适应地压缩 η. 在数值模拟上, 用 $\hat{\eta}_n$ 作为初始值 $v^{(0)}$ 有较好的完成.

2. **调节参数 α_0 和 ϖ 的选取**

在定理 7.4.4 中, 惩罚参数 $\alpha_r = \alpha_0|\hat{\eta}_{n,r}|^{-\varpi}$ 中的两个调节参数 α_0 和 ϖ 待定. 为了得到比较好的模拟效果, 可采用下面的 BIC 准则 (Chen et al., 2010)

$$f(\alpha_0,\varpi) = -\breve{\eta}^{\mathrm{T}}_{n(\alpha_0,\varpi)}\left(\frac{1}{n}\sum_{j=1}^{n}\hat{b}_{j,\hat{\theta}}\hat{b}^{\mathrm{T}}_{j,\hat{\theta}}\right)\breve{\eta}_{n(\alpha_0,\varpi)} + \frac{\log n}{n}(N_{(\alpha_0,\varpi)} - 1), \quad (7.4.16)$$

其中 $\breve{\eta}_{n(\alpha_0,\varpi)}$ 是在给定 (α_0,ϖ) 的情况下由式 (7.4.10) 得到的估计, $N_{(\alpha_0,\varpi)}$ 是 $\breve{\eta}_{n(\alpha_0,\varpi)}$ 中非零分量的个数, $\dfrac{\log n}{n}$ 是由 Li (2007) 建议的 BIC 惩罚因子.

通过对目标函数 (7.4.16) 极小化, 可得到 α_0 和 ϖ 的具体数值解, 求解可以通过两维空间格子点搜索的数值方法来实现. 具体而言, 预先给定 (α_0,ϖ) 的取值范围, 这个取值范围应尽可能的充裕, 使得极小化目标函数 (7.4.16) 得到的数值解应在这个取值范围的中间位置. 此外, 格子点个数应该设定在 20×20 左右.

7.4.6 模拟研究

本节通过数值模拟来考察在 7.4.3~7.4.5 节中提到的估计与变量选择方法. 在下面的两个例子中, 例 7.4.1 用于考察参数 (β,θ) 估计的效果. 例 7.4.2 用于考察 7.4.5 节提出关于 β 的变量选择方法的效果.

例 7.4.1 数据 $\{(\widetilde{Y}_i,\widetilde{X}_i,Z_i), i=1,\cdots,n\}$ 从下面的部分线性单指标模型产生

$$\begin{cases} Y = X^{\mathrm{T}}\beta + \sin(Z^{\mathrm{T}}\theta) + \varepsilon, \\ \widetilde{Y} = \phi(U)Y, \\ \widetilde{X}_r = \psi_r(U)X_r, \quad r = 1,\cdots,p, \end{cases} \quad (7.4.17)$$

其中$\beta = (3.0,1.5,1.5,0.5,0.5)^{\mathrm{T}}$, $\theta = (1,1,1)^{\mathrm{T}}/\sqrt{3}$, 误差 $\varepsilon \sim N(0,0.1^2)$ 并且与 (X,Z) 独立. 协变量 (X,Z) 服从多元正态分布 $N_8\left((\mu_X,\mu_Z)^{\mathrm{T}},\Sigma\right)$, 其中$\mu_X = (5,\cdots,5)^{\mathrm{T}}$, $\mu_Z = (0,0,0)^{\mathrm{T}}$, $\Sigma = \left(0.5^{|i-j|}\right)_{1\leqslant i,j\leqslant 8}$. 此外, 混淆变量 U 和 (X,Z,ε) 独立, 服从区间 [1, 6] 上的均匀分布. 失真函数选取为 $\phi(U) = \dfrac{(U+1)^2}{22.3333}$, $\psi_r(U) = \dfrac{U+r}{3.5+r}$, $r=1,\cdots,p$. 在本例中, $p=5$, 样本量分别取 $n = 100, 200, 300$ 和 400, 模拟的重复次数为 500.

在计算 $\hat{\theta}_{\mathrm{ini}}$ 的过程中, $H(\cdot)$ 选取标准多元正态分布密度函数, $L(t) = 0.75(1-t^2)I(|t|\leqslant 1)$. 这里采用了预先给定窗宽的方法: $h_1 = n^{-1/7}$ 和 $h_2 = n^{-1/5}$, 并且这

两个窗宽的选取满足定理 7.4.1 中的窗宽条件. 可验证, 即使 h_1 和 h_2 取上述给定窗宽的附近值时, 最终得到的 $\hat{\theta}_{\rm ini}$ 也表现比较稳定. 在节 7.4.3 中的迭代算法, 采用 Epanechnikov 核函数 $K(t) = 0.75(1-t^2)I(|t| \leqslant 1)$. 模拟结果报告在表 7.4.1 和表 7.4.2 中.

表 7.4.1 例 7.4.1 模型 (7.4.17) 中 θ 估计的平均偏差 (Bias) 和标准差 (SE)

n		$\hat{\theta}_1$	$\hat{\theta}_2$	$\hat{\theta}_3$	$\arccos(\hat{\theta},\theta)$
$n=100$	Bias	−0.0162	−0.0706	−0.0457	0.2848
	SE	0.1573	0.2672	0.2229	0.3393
$n=200$	Bias	0.0030	−0.0121	0.0013	0.0811
	SE	0.0547	0.0588	0.0465	0.0468
$n=300$	Bias	0.0027	−0.0051	−0.0004	0.0493
	SE	0.0318	0.0363	0.0288	0.0274
$n=400$	Bias	0.0008	−0.0013	−0.0013	0.0375
	SE	0.0240	0.0264	0.0233	0.0202

表 7.4.2 例 7.4.1 模型 (7.4.17) 中 β 估计的平均偏差 (Bias) 和标准差 (SE)

n		$\hat{\beta}_1$	$\hat{\beta}_2$	$\hat{\beta}_3$	$\hat{\beta}_4$	$\hat{\beta}_5$
$n=100$	Bias	0.0110	0.0006	−0.0014	−0.0085	0.0015
	SE	0.1094	0.0921	0.0910	0.0739	0.0832
$n=200$	Bias	−0.0003	−0.0012	−0.0087	−0.0043	−0.0044
	SE	0.0721	0.0480	0.0525	0.0419	0.0376
$n=300$	Bias	0.0075	−0.0013	−0.0023	−0.0006	−0.0021
	SE	0.0544	0.0390	0.0405	0.0274	0.0274
$n=400$	Bias	0.0090	0.0045	0.0004	−0.0010	−0.0011
	SE	0.0467	0.0344	0.0354	0.0216	0.0213

在表 7.4.1 和表 7.4.2 中列出了 θ 和 β 估计的平均偏差 (Bias) 和标准差 (SE). 从表中的模拟结果可以看出, 随着样本量的不断增加, $\hat{\beta}$ 和 $\hat{\theta}$ 的平均偏差和标准差不断减小, 且平均偏差也越来越趋于零. 此外, 在表 7.4.1 中还列出了 $\hat{\theta}$ 和真值 θ 的平均夹角 (以弧度表示, $\arccos\left(\hat{\theta},\theta\right)$) 和标准差, 并且随着样本量增加, 平均夹角越来越趋于零. 两个表格中的模拟结果都表明本节提出的估计方法表现很好.

例 7.4.2 本例考察变量选择方法的完成效果. 本例继续使用模型 (7.4.17), 重复模拟 500 次, 样本量选取为 $n = 300$, 400 和 500. 指标参数 θ 的选取和例 7.4.1 相同, 本例中 $\beta = (3,2,1.5,0.2,0.3,0.15,0,0,0,0,0,0)^{\rm T}$ 是 12 维的参数向量. 协变量 (X,Z) 取自多元正态分布 $N_{15}\left((\mu_X,\mu_Z)^{\rm T},\Sigma\right)$, 其中$\mu_X = (5,\cdots,5)^{\rm T}$ 和 $\mu_Z = (0,0,0)^{\rm T}$, $\Sigma = (0.5^{|i-j|})_{1\leqslant i,j\leqslant 15}$. 失真函数 $\phi(u)$ 和 $\psi_r(u)$ 的具体形式与例 7.4.1 相同. 此外, 本例对误差 ε 的设计考虑下面两种情形.

情形 1 ε 取自正态分布 $N(0,0.1^2)$, 且与 (X,Z) 独立.

情形 2 ε 取自正态分布 $N(0,0.1^2\times(|X_1|+|Z_1|))$, 其中 X_1 和 Z_1 分别是 X 和 Z 的第一个分量.

表 7.4.3 例 7.4.2 的变量选择模拟结果

n	$\omega_{u,\beta}(\%)$	$\omega_{c,\beta}(\%)$	$\omega_{o,\beta}(\%)$			Medse_β	No of zeros	
			"1" (%)	"2" (%)	"$\geqslant 3$"(%)		C_β	IN_β
				情形 1				
300	13.00	75.00	12.00	0.00	0.00	0.0133	5.840	0.160
400	0.20	85.80	13.80	0.20	0.00	0.0064	5.858	0.002
500	0.00	92.60	7.40	0.00	0.00	0.0047	5.926	0.000
				情形 2				
300	4.80	74.60	18.60	2.60	0.00	0.0112	5.746	0.050
400	0.20	84.40	14.00	1.20	0.20	0.0080	5.828	0.002
500	0.00	89.20	10.40	0.40	0.00	0.0049	5.888	0.000

调节参数 α_0 和 ϖ 的选取采用 7.4.5 节提出的 BIC 准则. 表 7.4.3 给出了变量选择的模拟结果. 表中 $\omega_{c,\beta}$ 表示恰好选对 6 个重要变量的情形, $\omega_{o,\beta}$ 表示多选变量的情形 (在把 6 个重要变量完全选对的情况下还多选出一些变量), $\omega_{u,\beta}$ 表示少选或错选变量的情形 ($\omega_{u,\beta}\%=100\%-\omega_{c,\beta}\%-\omega_{o,\beta}\%$). 同时, 在多选变量的情形中, 用符号 "1" "2" 和 "$\geqslant 3$" 表示多选 1 个, 多选 2 个和多选 3 个以上变量的情形. Medse_β 表示均方误差 $\|\hat{\beta}^s-\beta\|^2$ 的中位数, 此时 $\hat{\beta}^s$ 是指变量选择之后的估计, 见式 (7.4.11). 此外, 用符号 C_β 表示 β 中零系数被正确地估计成 0 的个数, IN_β 表示 β 中非零系数被错误地估计成 0 的个数.

从表 7.4.3 中的模拟结果可看出, 在两种误差分布的情形下, 第一稀疏主成分 $\breve{\eta}_n$ 能够正确的分辨出 β 中 0 元素与非 0 元素. C_β 和 IN_β 分别很接近理论真值 6 和 0. 总体上来说, 恰好选对 6 个变量的情况高达 70%以上, 多选变量约占 20%, 少选和选错约占 10%. 此外, 随着样本量的增大, 如 $n=500$, 恰好选对变量的情况可达 90%. 这表明本节提出变量选择方法在样本量比较大的时候效果较好. 此外, 从表 7.4.3 中也可看出, 中位数 Medse_β 随着样本量的增大递减到零的速度很快.

7.4.7 实际应用

本节分析一组糖尿病数据. 这组数据有 442 个观察病例, 其中响应变量 Y 取为在一年之后的糖尿病患者疾病进展的测量值, 协变量包括年龄 (AGE), 性别 (SEX), 体重指标 (BMI), 平均血压值 (BP) 和 6 个血清测量值有关于总胆固醇含量, 密度水平, 眼压性青光眼水平, 血糖浓度等, 分别记为 TC, LDL, HDL, TCH, LTG 和 GLU. 这组数据曾经被 Efron 等 (2004) 用最小角回归方法进行过系统分析. Efron 等 (2004) 用传统的线性模型来拟合这组数据并用最小角回归算法来进行选择与糖

尿病患者疾病有关的变量. 为了尽量避免模型失真, 这里采用部分线性单指标模型 (7.4.1) 来分析这组数据, 并考虑用 BMI 作为混淆变量 U, 用 AGE 和 BP 作为单指标部分的变量 Z, 而 SEX 以及 6 个血清测量值作为协变量 X.

这组数据中单指标部分估计为 $\hat{\theta} = (0.5010, -0.8655)^{\mathrm{T}}$, 即 0.5010AGE−0.8655BP. 第一稀疏主成分 (SPC) 的估计为

$$\breve{\eta}_n = (0.2502, 0.0344, -0.0366, -0.0144, 0.2327, -0.9384, 0.0033)^{\mathrm{T}}.$$

从 SPC 的估计来看, 所有 7 个变量 X 都应该保留在模型中, 进而 β 的估计为

$$\hat{\beta} = (-24.4754, 1.4202, -1.3503, -2.5827, -9.3805, 33.7842, 0.8867)^{\mathrm{T}}.$$

下面考察模型 (7.4.1) 中联系函数 $g(\cdot)$ 是否为线性函数. 基于 $\{(\widetilde{Y}_j - \widetilde{X}_j^{\mathrm{T}}\hat{b}_{j,\hat{\theta}}, Z_j^{\mathrm{T}}\hat{\theta}), j = 1, \cdots, n\}$, 可采用 Fan 和 Gijbels (1996) 提出的局部线性方法来估计 $g(\cdot)$, 并给出其 95%的置信带, 见图 7.4.1. 此外, 对数据 $\{\widetilde{Y}_j - \widetilde{X}_j^{\mathrm{T}}\hat{b}_{j,\hat{\theta}}, Z_j^{\mathrm{T}}\hat{\theta}\}_{j=1}^n$ 进行线

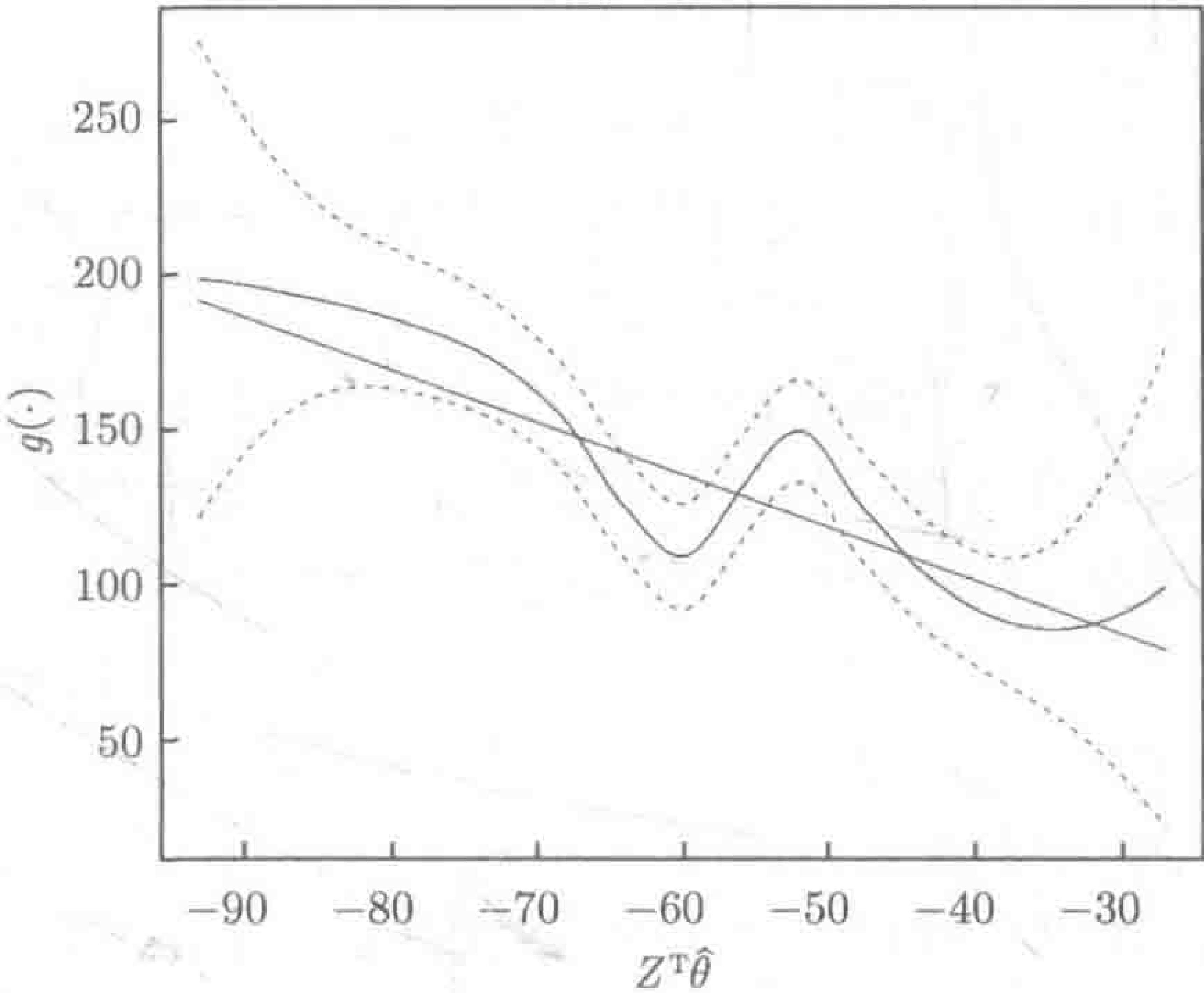


图 7.4.1 实曲线表示 $g(\cdot)$ 的局部线性拟合曲线, 虚线表示联系函数 $g(\cdot)$ 的置信水平为 95%的置信带, 直线是最小二乘拟合直线, 横轴表示指标 0.5010AGE−0.8655BP

性模型拟合. 从图 7.4.1 可以看出, 这条拟合直线并不在估计 $\hat{g}(\cdot)$ 的 95%置信带中, 因此说明联系函数 $g(\cdot)$ 确实是非线性的函数. Zhang 等 (2013b) 还采用了 Stute 和 Zhu (2005) 提出检验统计量的方法来检验联系函数 $g(\cdot)$ 对这组数据而言是否合适. 经过计算, 检验统计量的值为 1.4905, 其 p 值为 0.1361, 进一步说明联系函数 $g(\cdot)$ 是一个非线性函数.

7.5 部分线性单指标补救 EV 模型

7.5.1 模型介绍

本节用部分线性单指标模型来分析一类具有可加结构的测量误差数据. 具体而言, 线性部分 $X = (\xi^{\mathrm{T}}, W^{\mathrm{T}})^{\mathrm{T}}$ 中的变量 ξ 不能被观察, 但是可以得到一些辅助变量 (V, η) 来对不可观测的变量 ξ 进行补救. Zhang 等 (2014) 研究了如下部分线性补救 EV 模型:

$$\begin{cases} Y = \xi^{\mathrm{T}}\alpha + W^{\mathrm{T}}\beta + g(Z^{\mathrm{T}}\theta) + \varepsilon, \\ \eta = \xi(V) + e, \end{cases} \tag{7.5.1}$$

其中模型误差 e 满足 $E(e|V) = 0$, 并具有正定的协方差矩阵 $\Sigma_{ee} = E(ee^{\mathrm{T}})$. 此外, 模型 (7.5.2) 中待估的参数 $(\alpha^{\mathrm{T}}, \beta^{\mathrm{T}}) \in \mathbb{R}^r \times \mathbb{R}^p$, $r + p = d$. α 是不可观察变量 ξ 所对应的参数向量, β 是可观察变量 W 所对应的参数向量. 指标参数 $\theta \in \mathbb{R}^q$, 且满足 $\|\theta\| = 1$, $g(\cdot)$ 是未知的联系函数. 模型误差 ε 满足 $E[\varepsilon|\xi, W, Z] = 0$ 和 $E[\varepsilon^2|\xi, W, Z] < \infty$.

在这种情形下, 不可观测变量 ξ 可通过条件期望 $\xi(V) = E(\eta|V)$ 进行弥补. 这种类型的测量误差实际上包含了现有测量误差模型中的很多模型. Zhou 和 Liang (2009) 列举了一些具体的例子来说明模型 (7.5.2) 的应用价值, 其中包括 Cui 等 (2003) 和 Cui 和 Hu (2003) 所研究的降噪 (De-noise) 线性模型和降噪非线性模型. 经济中常用的理性预期等模型也可以看成是模型 (7.5.2) 的特例. 此外, 重复测量误差数据也可以看成模型 (7.5.2) 的一种特殊情形. 具体解释如下: 记 ξ 的两次独立重复观测值分别为 $S_1 = \xi + e_1$, $S_2 = \xi + e_2$. 因为是两次重复测量是独立进行的, 一个自然的假设就是 $E(e_1|S_2) = 0$, $E(e_2|S_1) = 0$. 因此有 $E(S_2|S_1 = s) = E(S_1|S_2 = s) = E(\xi|S = s)$. 那么可将 $\xi(s) = E(\xi|S = s)$ 的估计通过 $E(S_2|S_1 = s)$ 或者 $E(S_1|S_2 = s)$ 得出. 这也是模型 (7.5.2) 在处理重复测量误差这一实际问题时的一种应用.

Liang 和 Li (2009) 已经指出, 即使是最简单情形下的测量误差模型 ($W = X + U$), 如果忽略测量误差 U, 那么此时得到的最终估计可能是有偏估计, 同时也不可能选出正确的变量, 最终导致模型有偏. 本节将考虑部分线性单指标模型补救 EV 模型中参数 (α, β, θ) 的估计问题和变量选择问题.

7.5.2 估计方法和渐近性质

记 $\{(Y_i, \eta_i, V_i, W_i, Z_i), i = 1, \cdots, n\}$ 是来自于总体 (Y, η, V, W, Z) 的 i.i.d. 随机样本. 因 ξ_i 不可观测, 可先用其辅助变量 $\{(\eta_i, V_i), i = 1, \cdots, n\}$ 将 ξ_i 估计出来, 然后再利用估计出来的 $\hat{\xi}_i$ 和可观测数据 (Y_i, W_i, Z_i) 来给出参数 (α, β, θ) 的估计.

1. ξ 的估计

记 $\{(\eta_i, V_i), i=1,\cdots,n\}$ 是辅助变量 (η, V) 的 i.i.d. 随机样本. 首先介绍 ξ 的估计方法. 需指出的是, 虽然下面提出的估计方法可以很自然地推广到 V 是多维的情形, 但是多维情况下的"维数祸根"问题将导致参数估计的效率降低. 为此, 这里只考虑 V 的维数是 1 维的情形. 记 η_{ik} 是 η_i 的第 k 个分量, $i=1,\cdots,n$. 这里使用 Fan 和 Gijbels (1996) 提出的局部线性光滑方法将 $\xi(v)$ 的第 k 个分量 $\xi_k(v)$ 估计出来, 即关于 (a_{0k}, a_{1k}) 极小化下面的局部加权最小二乘目标函数:

$$\sum_{i=1}^{n}\left\{\eta_{ik}-a_{0k}-a_{1k}(V_i-v)\right\}^2 L_{b_k}(V_i-v), \tag{7.5.2}$$

其中 $L_{b_k}(\cdot)=L(\cdot/b_k)/b_k$, $L(\cdot)$ 是核密度函数, $b_k\ (k=1,\cdots,r)$ 是窗宽. 记 $(\hat{a}_{0k}, \hat{a}_{1k})$ 是从最小化目标函数 (7.5.2) 中得到的估计, 表示为

$$\hat{\xi}_k(v)=\hat{a}_{0k}=\frac{S_{20,k}(v)S_{01,k}(v)-S_{10,k}(v)S_{11,k}(v)}{S_{00,k}(v)S_{20,k}(v)-S_{10,k}^2(v)}, \tag{7.5.3}$$

其中 $S_{l_1l_2,k}(v)=\sum\limits_{i=1}^{n}L_{b_k}(V_i-v)(V_i-v)^{l_1}\eta_{ik}^{l_2}$, $l_1=0,1,2$, $l_2=0,1$, $k=1,\cdots,r$.

2. profile 最小二乘估计

在得到 ξ_i 的估计之后, 即 $\hat{\xi}_i=\left(\hat{\xi}_1(V_i),\cdots,\hat{\xi}_p(V_i)\right)^{\mathrm{T}}$, $i=1,\cdots,n$, 可用 profile 最小二乘来给出参数 (α,β,θ) 的估计. 先给定 $\zeta=(\alpha^{\mathrm{T}},\beta^{\mathrm{T}},\theta^{\mathrm{T}})^{\mathrm{T}}$, 同样采用局部线性光滑方法来给出联系函数 $g(u)$ 的估计. 极小化下面的局部加权最小二乘目标函数, 得到 (d_0,d_1) 的解:

$$\sum_{i=1}^{n}\left\{Y_i-\alpha^{\mathrm{T}}\hat{\xi}_i-\beta^{\mathrm{T}}W_i-d_0-d_1(Z^{\mathrm{T}}\theta_i-u)\right\}^2 K_h(Z^{\mathrm{T}}\theta_i-u), \tag{7.5.4}$$

其中 $K_h(\cdot)=K(\cdot/h)/h$, $K(\cdot)$ 是核密度函数, h 是窗宽. 此时有

$$\hat{g}(u,\zeta)=\hat{d}_0=\frac{T_{20}(u,\zeta)T_{01}(u,\zeta)-T_{10}(u,\zeta)T_{11}(u,\zeta)}{T_{00}(u,\zeta)T_{20}(u,\zeta)-T_{10}^2(u,\zeta)}, \tag{7.5.5}$$

其中

$$T_{l_1l_2}(u,\zeta)=\sum_{i=1}^{n}K_h(Z^{\mathrm{T}}\theta_i-u)(Z^{\mathrm{T}}\theta_i-u)^{l_1}\{Y_i-\alpha^{\mathrm{T}}\hat{\xi}_i-\beta^{\mathrm{T}}W_i\}^{l_2}, \quad l_1=0,1,2, l_2=0,1.$$

不失一般性, 假设 θ 的第一个分量为正数. 在对 $\zeta=(\alpha^{\mathrm{T}},\beta^{\mathrm{T}},\theta^{\mathrm{T}})^{\mathrm{T}}$ 进行估计之前, 同样地将指标参数 θ 由 $\mathbb{R}^q$ 中的球面转换为 $\mathbb{R}^{q-1}$ 中的球内, 即

$$\Theta=\left\{\left(\left(1-\sum_{s=2}^{q}\theta_s^2\right)^{1/2},\theta_2,\cdots,\theta_q\right)^{\mathrm{T}}\ \middle|\ \sum_{s=2}^{q}\theta_s^2<1\right\}.$$

这里重新定义 $\zeta_{(1)} = (\alpha^{\mathrm{T}}, \beta^{\mathrm{T}}, \theta_{(1)}^{\mathrm{T}})^{\mathrm{T}}$, 其中 $\theta_{(1)} = (\theta_2, \theta_3, \cdots, \theta_q)^{\mathrm{T}}$. 下面采用 "去一分量" profile 最小二乘方法来得到估计 $\hat{\zeta}_{(1)}$, 即关于 $\zeta_{(1)}$ 极小化下面的 profile 最小二乘目标函数:

$$Q(\zeta_{(1)}) = \sum_{i=1}^{n} \left\{ Y_i - \alpha^{\mathrm{T}} \hat{\xi}_i - \beta^{\mathrm{T}} W_i - \hat{g}(Z^{\mathrm{T}} \theta_i, \zeta) \right\}^2. \tag{7.5.6}$$

等价地, 通过解下面的方程来得到其估计:

$$\left. \frac{\partial Q(\zeta_{(1)})}{\partial \zeta_{(1)}} \right|_{\zeta_{(1)} = \hat{\zeta}_{(1)}} = 0. \tag{7.5.7}$$

可采用 Newton-Raphson 迭代的算法来求解方程 (7.5.7). 那么 ζ 的最终估计为

$$\hat{\zeta} = \hat{\zeta}(\hat{\zeta}_{(1)}) = \left(\hat{\alpha}^{\mathrm{T}}, \hat{\beta}^{\mathrm{T}}, \hat{\theta}^{\mathrm{T}} \right)^{\mathrm{T}},$$

其中 $\hat{\theta} = \left(\sqrt{1 - \|\hat{\theta}_{(1)}\|^2}, \hat{\theta}_{(1)}^{\mathrm{T}} \right)^{\mathrm{T}}$.

3. 渐近结果

定义

$$\begin{gathered} m_\xi(t) = E(\xi | Z^{\mathrm{T}} \theta = t), \quad m_W(t) = E(W | Z^{\mathrm{T}} \theta = t), \quad m_Z(Z | Z^{\mathrm{T}} \theta = t), \\ \widetilde{\xi} = \xi - m_\xi(Z^{\mathrm{T}} \theta), \quad \widetilde{W} = W - m_W(Z^{\mathrm{T}} \theta), \\ \widetilde{Z} = Z - m_Z(Z^{\mathrm{T}} \theta), \quad \widetilde{S}(v) = E(\widetilde{T} | V = v) \end{gathered}$$

其中 $\widetilde{T} = (\widetilde{\xi}^{\mathrm{T}}, \widetilde{W}^{\mathrm{T}}, g'(Z^{\mathrm{T}} \theta)(J_{\theta_{(1)}}^{\mathrm{T}} \widetilde{Z})^{\mathrm{T}})^{\mathrm{T}}$ 和

$$J_{\theta_{(1)}} = \begin{pmatrix} -\theta_{(1)}^{\mathrm{T}} \Big/ \sqrt{1 - \|\theta_{(1)}\|^2} \\ I_{r-1} \end{pmatrix}.$$

为了给出参数估计量的渐近结果, 下面列出了定理所需要的正则条件.

(C7.5.1) 函数 $g(t)$, $E(\xi | Z^{\mathrm{T}} \theta = t)$, $E(W | Z^{\mathrm{T}} \theta = t)$, $E(Z | Z^{\mathrm{T}} \theta = t)$ 和 $Z^{\mathrm{T}} \theta$ 的密度函数 $f_{Z^{\mathrm{T}} \theta}(t)$ 关于 t 都存在连续的二阶导数, 并且二阶导数在 $\mathcal{C} = \{t = Z^{\mathrm{T}} \theta : z \in \mathcal{Z} \subset \mathbb{R}^r, \theta \in \Theta\}$ 上满足一致 Lipschiz 连续条件, 其中 $\mathcal{Z}$ 是一个协变量 Z 取值的一个紧集. 此外还需要求密度函数在 $\mathcal{C}$ 上有大于零的下界.

(C7.5.2) 随机变量 V 的密度函数 $f_V(v)$ 在 $v \in \mathbb{V}$ 上有大于零的下界, 并且 $f_V(v)$ 和 $\widetilde{S}(v)$ 关于 v 有连续的二阶导数, 其二阶导数在 $\mathbb{V}$ 上需满足一致 Lipschiz 连续条件, 其中 $\mathbb{V}$ 是 V 取值的一个紧集.

(C7.5.3) $E[\xi^{\otimes 2} | Z^{\mathrm{T}} \theta = t]$, $E[W^{\otimes 2} | Z^{\mathrm{T}} \theta = t]$ 和 $E[Z^{\otimes 2} | Z^{\mathrm{T}} \theta = t]$ 在 $\mathcal{C}$ 上都是可逆矩阵.

(C7.5.4) 核函数 $L(\cdot)$ 和 $K(\cdot)$ 是连续对称的有界密度函数, 且满足 $\int t^2L(t)\mathrm{d}t \neq 0$, $\int t^2K(t)\mathrm{d}t \neq 0$ 和 $\int |t|^jL(t)\mathrm{d}t < \infty$, $\int |t|^jK(t)\mathrm{d}t < \infty$, $j = 1,2,3$. 此外, 还需 $K(\cdot)$ 的二阶导数有界.

(C7.5.5) 当 $n \to \infty$ 时, 窗宽 h 需满足 $\dfrac{(\log n)^{1+s}}{nh} \to 0$ 和 $nh^6 \to 0$, 其中 s 为某个大于零的常数.

(C7.5.6) 窗宽 $b_k, k = 1,\cdots,r$, 当 $n \to \infty$ 时, 满足 $nb_k^4 \to 0$ 和 $nb_k^2 \to \infty$.

(C7.5.7) 对所有的 $\lambda_{1,j}$, $\lambda_{2,k}$ 和 $\lambda_{3,s}$, $j = 1,\cdots,p$, $k = 1,\cdots,r$, $s = 1,\cdots,q-1$, 当 $n \to \infty$ 时, $\lambda_{1,j} \to 0$, $\sqrt{n}\lambda_{1,j} \to \infty$, $\lambda_{2,k} \to 0$, $\sqrt{n}\lambda_{2,k} \to \infty$ 和 $\lambda_{3,s} \to 0$, $\sqrt{n}\lambda_{3,s} \to \infty$. 此外还需

$$\lim_{n\to\infty}\inf \lim_{u\to 0+}\inf p'_{\lambda_{1,j}}(u)/\lambda_{1,j} > 0, \quad \lim_{n\to\infty}\inf \lim_{u\to 0+}\inf p'_{\lambda_{2,k}}(u)/\lambda_{2,k} > 0,$$
$$\lim_{n\to\infty}\inf \lim_{u\to 0+}\inf p'_{\lambda_{3,s}}(u)/\lambda_{3,s} > 0.$$

定理 7.5.1 假设条件(C7.5.1)~条件(C7.5.6)成立, 当 $n \to \infty$ 时, 有

$$\sqrt{n}\big(\hat{\zeta}_{(1)} - \zeta_{(1)}\big) \xrightarrow{L} N(0, \Sigma^*),$$

其中

$$\Sigma^* = \Sigma^{-1}D\Sigma^{-1}, \quad \Sigma = E\widetilde{T}^{\otimes 2}$$

$$D = E(\varepsilon\widetilde{T})^{\otimes 2} + E\big[\widetilde{S}(V)^{\otimes 2}(\alpha^{\mathrm{T}}e)^2\big] + E\big[\widetilde{T}\widetilde{S}(V)^{\mathrm{T}}(\varepsilon e^{\mathrm{T}})\big]\alpha + \alpha^{\mathrm{T}}E\big[(e\varepsilon)\widetilde{S}(V)\widetilde{T}^{\mathrm{T}}\big].$$

定理 7.5.1 的渐近方差中, $\Sigma^{-1}E(\varepsilon\widetilde{T})^{\otimes 2}\Sigma^{-1}$ 是没有测量误差 e 影响时的 profile 最小二乘估计的渐近方差, 其余项是由测量误差 e 和辅助变量 V 所引起. 此外, 如果假设在给定 (ξ, W, Z, V) 时, e 和 ε 独立, 同时假设 ε 和 (ξ, W, Z) 也是独立的, 那么定理 7.5.1 中的渐近方差可简化为 $\sigma^2\Sigma^{-1} + \alpha^{\mathrm{T}}\Sigma_{ee}\alpha\Sigma^{-1}\big[E\widetilde{S}(V)^{\otimes 2}\big]\Sigma^{-1}$.

推论 7.5.1 在定理 7.5.1 的条件下, 有

$$\sqrt{n}\big(\hat{\zeta} - \zeta\big) \xrightarrow{L} N\left(0, \Sigma_1\right),$$

其中 $\Sigma_1 = I_{(1)}\Sigma^* I_{(1)}^{\mathrm{T}}$, $I_{(1)} = \mathrm{diag}(I_p, I_q, J_{\theta_{(1)}})$.

这个推论可基于定理 7.5.1 结论, 直接使用多元 Delta 方法得到. 如果 $e = 0$ 并且 $\xi(v) = v$, 那么推论 7.5.1 的渐近方差和 Cui 等 (2011) 中推论 1 的渐近方差类似.

7.5.3　变量选择

1. 变量选择程序

本节采用 Liang 等 (2010) 提出的惩罚 profile 最小二乘方法来达到变量选择的目的. 但是这里在惩罚指标参数 θ 的时候与 Liang 等 (2010) 采用的措施不同, Zhang 等 (2014) 提出如下的惩罚目标函数

$$\mathcal{L}_P(\zeta_{(1)}) = \frac{1}{2}Q(\zeta_{(1)}) + n\sum_{j=1}^{r} p_{\lambda_{1,j}}(|\alpha_j|) + n\sum_{j=1}^{p} p_{\lambda_{2,j}}(|\beta_j|) + n\sum_{j=1}^{q-1} p_{\lambda_{3,j}}(|\theta_{(1),j}|), \quad (7.5.8)$$

其中 $p_\lambda(\cdot)$ 是一个惩罚函数, $\lambda_{1,j}$, $\lambda_{2,k}$ 和 $\lambda_{3,s}$ 是截断参数. 上述惩罚目标函数是对 $q-1$ 维参数向量 $\theta_{(1)}$ 进行惩罚, 这与 Liang 等 (2010) 中使用的对 q 维指标参数直接进行惩罚选变量的方法不同. 因为根据指标参数的识别性条件, θ 的参数空间被限定为 $\{\theta : \|\theta\| = 1, \theta_1 > 0\}$. 也就是说, 指标参数 θ 的第一个分量 θ_1 是正数, 同时也表明协变量 Z 的第一个分量 Z_1 应该在模型当中, 而不需要通过变量选择的方法来选择得到 Z_1. 基于此事实, 可将惩罚函数 $p_\lambda(\cdot)$ 加在 $\theta_{(1)}$ 上来选择剩变量 $(Z_2, \cdots, Z_q)$ 中哪些变量与模型 (7.5.2) 有关. 本节提出了 “去一分量”profile 惩罚最小二乘变量选择方法.

目前流行的两种惩罚函数是由 Tibshirani (1996) 提出的 L_1 惩罚函数 Lasso 与 Fan 和 Li (2001) 提出的 SCAD 惩罚. 关于惩罚变量选择方法的更多讨论, 见本书 1.5 节. 下面引入 SCAD 惩罚函数 $p_\lambda(\cdot)$, 该函数满足 $p_\lambda(0) = 0$ 以及其 1 阶导数满足

$$p'_\lambda(\gamma) = \lambda\Big\{I(\gamma \leqslant \lambda) + \frac{(a\lambda-\gamma)_+}{(a-1)\lambda}I(\gamma > \lambda)\Big\}.$$

Fan 和 Li (2001) 建议在实际使用中常数 a 取 3.7. 本节中的模拟试验中也继续采用该数值.

2. 渐近性质

接下来给出变量选择方法的大样本理论结果. 不失一般性, 假设 $\alpha = (\alpha_{\mathbf{1}}^{\mathrm{T}}, \alpha_{\mathbf{2}}^{\mathrm{T}})^{\mathrm{T}}$, $\beta = (\beta_{\mathbf{1}}^{\mathrm{T}}, \beta_{\mathbf{2}}^{\mathrm{T}})^{\mathrm{T}}$ 和 $\theta_{(1)} = \left(\theta_{\mathbf{1}(1)}^{\mathrm{T}}, \theta_{\mathbf{2}(1)}^{\mathrm{T}}\right)^{\mathrm{T}}$, 其中 $\alpha_{\mathbf{1}}$, $\beta_{\mathbf{1}}$ 和 $\theta_{\mathbf{1}(1)}$ 分别是 $r_0 \times 1$, $p_0 \times 1$ 和 $(q_0 - 1) \times 1$ 的非零分量, 而剩余的 $\alpha_{\mathbf{2}}$, $\beta_{\mathbf{2}}$ 和 $\theta_{\mathbf{2}(1)}$ 是 $(r - r_0) \times 1$, $(p - p_0) \times 1$ 和 $(q - q_0) \times 1$ 的零向量. 记 $\theta_{\mathbf{1}*} = (\theta_1, \theta_{\mathbf{1}(1)}^{\mathrm{T}})^{\mathrm{T}}$. 另外, $\xi_{\mathbf{1}}$, $W_{\mathbf{1}}$, $Z_{\mathbf{1}}$ 和 $e_{\mathbf{1}}$ 是对应于 ξ, W, Z 和 e 与模型 (7.5.2) 有关的分量.

定义

$$\widetilde{\xi_{\mathbf{1}}} = \xi_{\mathbf{1}} - E\left(\xi_{\mathbf{1}}|\theta_{\mathbf{1}*}^{\mathrm{T}} Z_{\mathbf{1}}\right), \quad \widetilde{W_{\mathbf{1}}} = W_{\mathbf{1}} - E\left(W_{\mathbf{1}}|\theta_{\mathbf{1}*}^{\mathrm{T}} Z_{\mathbf{1}}\right),$$
$$\widetilde{Z}_{\mathbf{1}} = Z_{\mathbf{1}} - E\left(Z_{\mathbf{1}}|\theta_{\mathbf{1}*}^{\mathrm{T}} Z_{\mathbf{1}}\right).$$

此外, $\widetilde{T}_{\mathbf{1}}=(\widetilde{\xi}_{\mathbf{1}}^{\mathrm{T}},\widetilde{W}_{\mathbf{1}}^{\mathrm{T}},g'(\theta_{\mathbf{1}*}^{\mathrm{T}}Z_{\mathbf{1}})(J_{\theta_{\mathbf{1}(1)}}^{\mathrm{T}}\widetilde{Z}_{\mathbf{1}})^{\mathrm{T}})^{\mathrm{T}}$, $\widetilde{S}_{\mathbf{1}}(v)=E(\widetilde{T}_{\mathbf{1}}|V=v)$, $\Gamma_{\mathbf{1}}=E\widetilde{T}_{\mathbf{1}}^{\otimes 2}$,

$$J_{\theta_{\mathbf{1}(1)}}=\begin{pmatrix}-\theta_{\mathbf{1}(1)}^{\mathrm{T}}\Big/\sqrt{1-\|\theta_{\mathbf{1}(1)}\|^2}\\ I_{q_0-1}\end{pmatrix}.$$

进一步, 定义下面的记号

$$\begin{aligned}\mathcal{R}_{\lambda_{\mathbf{1}}}&=\Big\{p'_{\lambda_{1,1}}(|\alpha_{\mathbf{1},1}|)\mathrm{sgn}(\alpha_{\mathbf{1},1}),\cdots,p'_{\lambda_{1,r_0}}(|\alpha_{\mathbf{1},r_0}|)\mathrm{sgn}(\alpha_{\mathbf{1},r_0})\Big\},\\ \mathcal{R}_{\lambda_2}&=\Big\{p'_{\lambda_{2,1}}(|\beta_{\mathbf{1},1}|)\mathrm{sgn}(\beta_{\mathbf{1},1}),\cdots,p'_{\lambda_{2,p_0}}(|\beta_{\mathbf{1},p_0}|)\mathrm{sgn}(\beta_{\mathbf{1},p_0})\Big\},\\ \mathcal{R}_{\lambda_3}&=\Big\{p'_{\lambda_{3,1}}(|\theta_{\mathbf{1}(1),1}|)\mathrm{sgn}(\theta_{\mathbf{1}(1),1}),\cdots,p'_{\lambda_{3,q_0-1}}(|\theta_{\mathbf{1}(1),q_0-1}|)\mathrm{sgn}(\theta_{\mathbf{1}(1),q_0-1})\Big\}\end{aligned}$$

和

$$\begin{aligned}\Sigma_{\lambda_1}&=\mathrm{diag}\Big\{p''_{\lambda_{1,1}}(|\alpha_{\mathbf{1},1}|),\cdots,p''_{\lambda_{1,r_0}}(|\alpha_{\mathbf{1},r_0}|)\Big\},\\ \Sigma_{\lambda_2}&=\mathrm{diag}\Big\{p''_{\lambda_{2,1}}(|\beta_{\mathbf{1},1}|),\cdots,p''_{\lambda_{2,p_0}}(|\beta_{\mathbf{1},p_0}|)\Big\},\\ \Sigma_{\lambda_3}&=\mathrm{diag}\Big\{p''_{\lambda_{3,1}}(|\theta_{\mathbf{1}(1),1}|),\cdots,p''_{\lambda_{3,q_0-1}}(|\theta_{\mathbf{1}(1),q_0-1}|)\Big\},\end{aligned}$$

其中$\alpha_{\mathbf{1},j}$, $\beta_{\mathbf{1},k}$ 和 $\theta_{\mathbf{1}(1),s}$ 分别是 $\alpha_{\mathbf{1}}$, $\beta_{\mathbf{1}}$ 和 $\theta_{\mathbf{1}(1)}$ 的第 j 个, 第 k 个和第 s 个分量. 再记 $\mathcal{R}_\lambda=(\mathcal{R}_{\lambda_1}^{\mathrm{T}},\mathcal{R}_{\lambda_2}^{\mathrm{T}},\mathcal{R}_{\lambda_3}^{\mathrm{T}})^{\mathrm{T}}$ 和 $\Sigma_\lambda=\mathrm{diag}\,(\Sigma_{\lambda_1},\Sigma_{\lambda_2},\Sigma_{\lambda_3})$.

定义由目标函数 (7.5.8) 得到的惩罚 profile 最小二乘估计为 $\hat{\alpha}_{\lambda_1}=(\hat{\alpha}_{\lambda_1\mathbf{1}}^{\mathrm{T}},\hat{\alpha}_{\lambda_1\mathbf{2}}^{\mathrm{T}})^{\mathrm{T}}$, $\hat{\beta}_{\lambda_2}=(\hat{\beta}_{\lambda_2\mathbf{1}}^{\mathrm{T}},\hat{\beta}_{\lambda_2\mathbf{2}}^{\mathrm{T}})^{\mathrm{T}}$ 和 $\hat{\theta}_{\lambda_3(1)}=(\hat{\theta}_{\lambda_3\mathbf{1}(1)}^{\mathrm{T}},\hat{\theta}_{\lambda_3\mathbf{2}(1)}^{\mathrm{T}})^{\mathrm{T}}$.

定理 7.5.2 假设条件(C7.5.1)~条件(C7.5.7)成立, 当 $n\to\infty$ 时, 有

(1) $\hat{\alpha}_{\lambda_1\mathbf{2}}=0$, $\hat{\beta}_{\lambda_2\mathbf{2}}=0$, $\hat{\theta}_{\lambda_3\mathbf{2}(1)}=0$ 依概率 1 成立;

(2) 记 $\hat{\zeta}_{\lambda\mathbf{1}(1)}=\left(\hat{\alpha}_{\lambda_1\mathbf{1}}^{\mathrm{T}},\hat{\beta}_{\lambda_2\mathbf{1}}^{\mathrm{T}},\hat{\theta}_{\lambda_3\mathbf{1}(1)}^{\mathrm{T}}\right)^{\mathrm{T}}$, $\zeta_{\mathbf{1}(1)}=(\alpha_{\mathbf{1}}^{\mathrm{T}},\beta_{\mathbf{1}}^{\mathrm{T}},\theta_{\mathbf{1}(1)}^{\mathrm{T}})^{\mathrm{T}}$, 有

$$\sqrt{n}\,(\Gamma_{\mathbf{1}}+\Sigma_\lambda)\left\{\left(\hat{\zeta}_{\lambda\mathbf{1}(1)}-\zeta_{\mathbf{1}(1)}\right)-(\Gamma_{\mathbf{1}}+\Sigma_\lambda)^{-1}\mathcal{R}_\lambda\right\}\xrightarrow{L}N(0,D_{\mathbf{1}}),$$

其中渐近方差矩阵定义为

$$\begin{aligned}D_{\mathbf{1}}=&E(\varepsilon\widetilde{T}_{\mathbf{1}})^{\otimes 2}+E\big[\widetilde{S}_{\mathbf{1}}(V)^{\otimes 2}(\alpha_{\mathbf{1}}^{\mathrm{T}}e_{\mathbf{1}})^2\big]+E\big[\widetilde{T}_{\mathbf{1}}\widetilde{S}_{\mathbf{1}}(V)^{\mathrm{T}}(\varepsilon e_{\mathbf{1}}^{\mathrm{T}})\big]\alpha_{\mathbf{1}}\\ &+\alpha_{\mathbf{1}}^{\mathrm{T}}E\big[(e_{\mathbf{1}}\varepsilon)\widetilde{S}_{\mathbf{1}}(V)\widetilde{T}_{\mathbf{1}}^{\mathrm{T}}\big].\end{aligned}$$

这里值得指出的是本节将指标参数 θ 从 $\mathbb{R}^q$ 的球面转化为 $\mathbb{R}^{q-1}$ 的球内之后, 不需要对 θ 的第一个分量进行惩罚, 而只需对其余的 $q-1$ 个分量进行惩罚, 这样可提高计算效率. 定理 7.5.2(2) 说明当选择合适的截断参数 $\lambda_{1,j}$, $\lambda_{2,k}$, $\lambda_{3,s}$ 之后, SCAD 变量选择方法可保持变量选择的 oracle 性质.

记$\hat{\zeta}_{\lambda 1}=\left(\hat{\alpha}_{\lambda_1 1}^{\mathrm{T}},\hat{\beta}_{\lambda_2 1}^{\mathrm{T}},\hat{\theta}_{\lambda_3 1}^{\mathrm{T}}\right)^{\mathrm{T}}$, $\zeta_1=(\alpha_1^{\mathrm{T}},\beta_1^{\mathrm{T}},\theta_{1*}^{\mathrm{T}})^{\mathrm{T}}$, 其中 $\hat{\theta}_{\lambda_3 1}=\left(\hat{\theta}_{\lambda_3 1},\hat{\theta}_{\lambda_3 1(1)}^{\mathrm{T}}\right)^{\mathrm{T}}$, $\hat{\theta}_{\lambda_3 1}=\sqrt{1-\|\hat{\theta}_{\lambda_3 1(1)}\|^2}$. 从定理 7.5.2, 可以得到下面的推论.

推论 7.5.2 在定理 7.5.2 条件下, 如果 $\sqrt{n}\mathcal{R}_\lambda=o_P(1)$, $\Sigma_\lambda=o_P(1)$, 有

$$\sqrt{n}\left(\hat{\zeta}_{\lambda 1}-\zeta_1\right)\xrightarrow{L}N\left(0,\Sigma_1\right),$$

其中 $\Sigma_1=I_1\Gamma_1 I_1^{\mathrm{T}}$, $\Gamma_1=\Gamma_1^{-1}D_1\Gamma_1^{-1}$ 和 $I_1=\operatorname{diag}(I_p,I_q,J_{\theta_{1(1)}})$.

3. 截断参数的选取

Liang 等 (2010) 建议使用 Wang 等 (2009) 提出的 BIC 准则来选择截断参数. 为了避免在高维空间 $(r+p+q-1)$ 搜索最优的 $\lambda_{1,j}$, $\lambda_{2,k}$ 和 $\lambda_{3,s}$. Liang 等 (2010) 提出了下面方法: $\lambda_{1,j}=\lambda_0\mathrm{SE}(\hat{\alpha}_j)$, $\lambda_{2k}=\lambda_0\mathrm{SE}(\hat{\beta}_k)$, $\lambda_{3,s}=\lambda_0\mathrm{SE}(\hat{\theta}_{(1)s})$, 其中 $\mathrm{SE}(\hat{\alpha}_j)$, $\mathrm{SE}(\hat{\beta}_k)$ 和 $\mathrm{SE}(\hat{\theta}_{(1)k})$ 分别是 α_j, β_k 和 $\theta_{(1)s}$, $j=1,\cdots,r$, $k=1,\cdots,p$, $s=1,\cdots,q-1$, 对应的标准差.

定义下面的 BIC 得分函数

$$\mathrm{BIC}(\lambda_0)=\log\{\mathrm{MSE}(\lambda_0)\}+\frac{\log n}{n}N_{\lambda_0}, \tag{7.5.9}$$

其中$\mathrm{MSE}(\lambda_0)=\dfrac{1}{n}\sum\limits_{i=1}^{n}\{Y_i-\hat{\alpha}_{\lambda_1}^{\mathrm{T}}\hat{\xi}_i-\hat{\beta}_{\lambda_2}^{\mathrm{T}}W_i-\hat{g}(\hat{\theta}_{\lambda_3}^{\mathrm{T}}Z_i)\}^2$, N_{λ_0} 是 $\hat{\alpha}_{\lambda_1}$,$\hat{\beta}_{\lambda_2}$,$\hat{\theta}_{\lambda_3}$ 中非零元素的个数. 这样关于 $\lambda_{1,j}$,$\lambda_{2,k}$ 和 $\lambda_{3,s}$ 的搜索就可以转化为对 λ_0 在 1 维空间中的搜索.

在上述 BIC 准则的定义中, 可将 $\log n$ 替换为 $2(r+p+q-1)$, 这样替换之后相应的准则也被称为 AIC 准则 (Akaike, 1973). 模拟研究中将对比 SCAD-AIC 和 SCAD-BIC 这两种法则, 看哪种准则会得到更好的变量选择效果.

7.5.4 模拟研究

本节用数值模拟来说明估计方法与变量选择方法的完成情况.

例 7.5.1 从下面的模型产生样本量为 200 的数据集, 重复 500 次试验.

$$\begin{cases}Y=\alpha^{\mathrm{T}}\xi+\beta^{\mathrm{T}}W+3\sin\left(Z^{\mathrm{T}}\theta\right)+\varepsilon,\\ \xi=\xi(V),\\ \eta=\xi(V)+e,\end{cases} \tag{7.5.10}$$

其中$\alpha=(1,3,2.5)^{\mathrm{T}}$, $\beta=(2,-1.5,0.5)^{\mathrm{T}}$ 和 $\theta=(1,1,1)^{\mathrm{T}}/\sqrt{3}$. (W,Z) 来自于均值为 0, 协方差矩阵为 $\Sigma=\left(0.5^{|i-j|}\right)_{1\leqslant i,j\leqslant 6}$ 的正态分布. 这里考虑不可观测的协变量 ξ 与

其可观侧的辅助变量 (η, V) 有如下的关系成立: $\xi=(\xi_1,\xi_2,\xi_3)^{\mathrm{T}}$, $\xi_1(V)=V(1-V)$, $\xi_2(V)=3\sin(V)$, $\xi_3(V)=3V-2\cos(V)$ 和 $\eta=\xi(V)+e$. 协变量 V 来自均值为 0, 方差为 0.5^2 的正态分布. 对于误差 ε 和 e, 考察下面两种情形.

情形 1 ε 取自于正态分布 $N(0,\sigma_\varepsilon^2)$, 并且与 (ξ, W, Z) 独立. e 和 (V,W,Z,ε) 独立, 产生于正态分布 $N\left(0,\sigma_e^2\Sigma\right)$, $\Sigma=\left(0.5^{|i-j|}\right)_{1\leqslant i,j\leqslant 3}$, 并考虑取各自的方差为 $\sigma_\varepsilon=0.1, \sigma_e=0.1$ 和 $\sigma_\varepsilon=0.5, \sigma_e=0.5$.

情形 2 ε 取自于正态分布 $N\big(0,\sigma_\varepsilon^2(|\xi_1(V)|+|W_1|+|Z_1|)\big)$, 其中 W_1 和 Z_1 分别是 W 和 Z 的第一个分量. e 来自正态分布 $N\left(0,\sigma_e^2V^2\Sigma\right)$. 这里的 $(\sigma_\varepsilon,\sigma_e,\Sigma)$ 的设计与情形 1 相同.

在估计 $\xi(v)$ 和 $g(\cdot)$ 的过程中, 使用 Epanechnikov 核函数 $L(t)=K(t)=0.75(1-t^2)I(|t|\leqslant 1)$. 在定理 7.5.1 的条件中, 需要对 $\xi(v)$ 采用欠光滑窗宽 b_k. 为了解决该问题, Zhou 和 Liang (2009) 给出了一个拇指法则: 在估计 $\xi(v)$ 的时候, 其窗宽可采用 $\hat{\sigma}_V n^{-1/3}$, 其中 $\hat{\sigma}_V$ 是 V 的样本标准差. 对于窗宽 h 的选择, 可使用去一分量的交错核实方法, 可见 Fan 和 Gijbels (1996). 模拟结果报告在表 7.5.1 中, 表中给出了三种估计量的模拟效果. 第一种是基准估计量 (benchmark estimator), 即 ξ 可以被准确地观测到; 第二种是本节提出的使用辅助变量 (η, V) 估计 ξ 的方法; 第三种是直接使用 η 而忽略误差 e 的估计方法. 表 7.5.1 中 "MEAN" 表示均值, "SD" 表示标准差, "B" 表示使用 ξ 的方法, "P" 表示使用 $\hat{\xi}$ 的方法, "N" 表示直接使用 η 而忽略 e 的方法.

从表 7.5.1 中可以看出, 使用辅助变量的估计方法和基准估计量都能很好地给出参数 (α,β,θ) 的估计. 然而, 直接使用 η 而忽略误差 e 之后, 得到的估计将导致很大的偏差. 表 7.5.1 中的模拟结果再次验证了这一事实.

此外, 对指标参数也计算了估计 $\hat{\theta}$ 和真值 θ 的平均夹角. 随着样本量的增加, 基准估计量和使用辅助变量估计的这两种方法得到的平均夹角都接近于 0, 再次说明使用辅助变量估计方法的可行性. 同时也可以看出, 当误差方差增加到 $\sigma_\varepsilon=0.5$ 和 $\sigma_e=0.5$ 时, 直接使用 η 而忽略 e 的估计, 其平均夹角已经大于 0.20. 说明忽略误差 e 时, 偏差不会消失, 此时得到的估计是不相合的.

例 7.5.2 本例考察变量选择方法的完成情况. 继续考虑模型 (7.5.10), 模拟 300 次, 每次的样本量为 $n=200$. 参数取为$\alpha=(2,0.5,0)^{\mathrm{T}}$, $\beta=(2,-1.5,1,0,0,0)^{\mathrm{T}}$ 和 $\theta=\left(\frac{1}{\sqrt{3}},\frac{1}{\sqrt{3}},\frac{1}{\sqrt{3}},0,0,0\right)^{\mathrm{T}}$. (W,Z) 来自于均值为 0, 协方差为 $\Sigma_{W,Z}=(0.5^{|i-j|})_{1\leqslant i,j\leqslant 12}$ 的正态分布. 不可观察的 ξ 与其辅助变量 (η,V) 之间建立如下关系: $\xi_1(V)=V(1-V)+2$, $\xi_2(V)=3\cos(V)-V$, $\xi_3(V)=-2\sin(V)$ 和 $\eta=\xi(V)+e$. 在本例中, V 来自均匀分布 $[0,1]$. 误差 (ε,e) 的分布与例 7.5.1 相同, 但是方差分别为 $\sigma_\varepsilon=0.1,\sigma_e=0.25$ 和 $\sigma_\varepsilon=0.5,\sigma_e=1$.

表 7.5.1　例 7.5.1 模拟结果

方法		α_{01}	α_{02}	α_{03}	β_{01}	β_{02}	β_{03}	θ_{01}	θ_{02}	θ_{03}	$\arccos(\hat{\theta},\theta)$
					情形 1: $\sigma_\varepsilon=0.1,\sigma_e=0.1$						
B	MEAN	0.9562	2.9228	2.5545	2.0003	−1.4990	0.4870	0.5772	0.5765	0.5777	0.0105
	SD	0.2379	0.3733	0.2589	0.0199	0.0235	0.0317	0.0224	0.0076	0.0149	0.0261
P	MEAN	1.1727	3.0227	2.5914	2.0001	−1.4991	0.4870	0.5778	0.5768	0.5769	0.0118
	SD	0.3735	0.4168	0.3506	0.0223	0.0261	0.0348	0.0173	0.0091	0.0114	0.0195
N	MEAN	0.4435	2.4212	2.8136	2.0009	−1.4993	0.4944	0.5637	0.5760	0.5881	0.0522
	SD	0.2961	0.4391	0.3128	0.0498	0.0561	0.0629	0.0471	0.0327	0.0354	0.0461
					情形 1: $\sigma_\varepsilon=0.5,\sigma_e=0.5$						
B	MEAN	0.9321	2.8950	2.5723	1.9999	−1.5030	0.4978	0.5641	0.5772	0.5865	0.0526
	SD	0.3094	0.4770	0.3310	0.0458	0.0537	0.0619	0.0463	0.0328	0.0393	0.0477
P	MEAN	1.0599	2.8685	2.6894	2.0007	−1.5039	0.5029	0.5606	0.5758	0.5903	0.0627
	SD	0.6948	0.7144	0.4814	0.0505	0.0578	0.0709	0.0491	0.0386	0.0432	0.0479
N	MEAN	−1.2796	1.5393	2.7647	1.9976	−1.5099	0.6555	0.3284	0.5572	0.7519	0.3227
	SD	0.2864	0.3685	0.3469	0.1854	0.2028	0.1991	0.0836	0.0770	0.0596	0.0800

续表

方法		α_{01}	α_{02}	α_{03}	β_{01}	β_{02}	β_{03}	θ_{01}	θ_{02}	θ_{03}	$\arccos(\hat{\theta},\theta)$
情形 2: $\sigma_\varepsilon=0.1,\sigma_e=0.1$											
B	MEAN	0.9578	2.9249	2.5537	1.9995	−1.5012	0.4885	0.5787	0.5763	0.5769	0.0114
	SD	0.2344	0.3671	0.2547	0.0218	0.0246	0.0306	0.0078	0.0085	0.0083	0.0086
P	MEAN	1.1150	3.0002	2.5857	1.9986	−1.4991	0.4909	0.5759	0.5769	0.5781	0.0172
	SD	0.4691	0.5689	0.3869	0.0341	0.0384	0.0430	0.0226	0.0173	0.0215	0.0314
N	MEAN	0.8027	2.7689	2.6367	2.0000	−1.5006	0.4919	0.5754	0.5753	0.5800	0.0265
	SD	0.2822	0.4051	0.2820	0.0319	0.0371	0.0452	0.0249	0.0198	0.0241	0.0302
情形 2: $\sigma_\varepsilon=0.5,\sigma_e=0.5$											
B	MEAN	0.9492	2.9101	2.5635	2.0011	−1.5007	0.5063	0.5595	0.5741	0.5916	0.0709
	SD	0.4384	0.6542	0.4544	0.0713	0.0750	0.0820	0.0535	0.0466	0.0479	0.0533
P	MEAN	0.9678	2.8196	2.6955	2.0021	−1.5011	0.5135	0.5490	0.5746	0.5982	0.0857
	SD	0.8100	0.8849	0.6653	0.0773	0.0820	0.0914	0.0657	0.0534	0.0580	0.0669
N	MEAN	−0.5926	1.5521	3.1812	2.0010	−1.5036	0.5747	0.4634	0.5731	0.6573	0.1846
	SD	0.4229	0.5600	0.4478	0.1175	0.1331	0.1403	0.1072	0.0782	0.0845	0.1018

模拟研究中比较了 SCAD-AIC 和 SCAD-BIC 两种截断参数选择法则方法, 看哪种准则会得到更好的变量选择效果. 跟例 7.5.1 讨论的情形相同, 本例仍旧考察三种情形: 基准统计量方法, 使用辅助变量估计 ξ 方法和忽略 e 而直接使用 η 的方法. 模拟结果见表 7.5.2, 表中 "C" 表示参数中 0 分量被正确压缩成 0 的平均数目, "I" 表示参数中非 0 分量被错误的压缩成 0 的平均数目, "oracle" 表示真实 0 与非 0 的个数, "S-AIC" 表示 SCAD-AIC 方法, "S-BIC" 表示 SCAD-BIC 方法, "B" 表示使用 ξ 的方法, "P" 表示使用 $\hat{\xi}$ 的方法, "N" 表示直接使用 η 而忽略 e 的方法.

表 7.5.2 例 7.5.2 模拟结果

方法		α		β		θ	
		C	I	C	I	C	I
情形 1: $\sigma_\varepsilon=0.1,\sigma_e=0.25$							
B	oracle	1	0	3	0	3	0
	S-AIC	0.4067	0	2.8867	0	2.9333	0
	S-BIC	0.5367	0	2.9833	0	2.9833	0
P	S-AIC	0.3500	0.0367	2.8867	0	2.9233	0
	S-BIC	0.4367	0.0733	2.9667	0.0100	2.9867	0
N	S-AIC	0.4367	0.3200	2.9933	0	2.9900	0
	S-BIC	0.4800	0.3367	2.9967	0.0167	2.9900	0
情形 1: $\sigma_\varepsilon=0.5,\sigma_e=1$							
B	oracle	1	0	3	0	3	0
	S-AIC	0.8100	0.0867	2.9367	0	2.6267	0
	S-BIC	0.8800	0.1000	2.9733	0	2.7867	0
P	S-AIC	0.3600	0.1867	2.9433	0	2.5900	0
	S-BIC	0.4433	0.2433	2.9733	0	2.8033	0
N	S-AIC	0.4933	0.9533	2.9800	0	2.8233	0
	S-BIC	0.5700	0.9767	2.9867	0	2.8900	0
情形 2: $\sigma_\varepsilon=0.1,\sigma_e=0.25$							
B	oracle	1	0	3	0	3	0
	S-AIC	0.4400	0.0033	2.8533	0	2.7833	0
	S-BIC	0.6267	0.0067	2.9733	0	2.9533	0
P	S-AIC	0.3767	0.0500	2.8633	0	2.7767	0
	S-BIC	0.5233	0.1133	2.9700	0.0100	2.9333	0
N	S-AIC	0.4433	0.1767	2.9467	0	2.8600	0
	S-BIC	0.5300	0.1900	2.9767	0	2.9233	0
情形 2: $\sigma_\varepsilon=0.5,\sigma_e=1$							
B	oracle	1	0	3	0	3	0
	S-AIC	0.6467	0.3233	2.9600	0	2.2000	0
	S-BIC	0.6500	0.3500	2.9633	0	2.2533	0
P	S-AIC	0.6033	0.3900	2.9633	0	2.2367	0
	S-BIC	0.6133	0.4367	2.9700	0	2.3267	0
N	S-AIC	0.6667	0.8133	2.9933	0	2.4567	0
	S-BIC	0.7500	0.8467	2.9967	0	2.5367	0

从表 7.5.2 中的模拟结果可以看出:

(1) SCAD-BIC 比 SCAD-AIC 的模拟效果要优越很多, SCAD-BIC 可以更加准确地将模型参数中的零分量挑选出来;

(2) 忽略误差 e 直接使用 η 所带来的后果, 该方法会将 α 中的非零分量错误的压缩成 0, 特别是当误差方差增加到 $\sigma_\varepsilon = 0.5$ 和 $\sigma_e = 1$ 时, SCAD-AIC 和 SCAD-BIC 几乎在这 300 个数据集中每次都将 α 中的某个非 0 分量错误地估计为 0, 这也说明, 如果直接使用 η 而忽略 e 时, SCAD-AIC 和 SCAD-BIC 这两种方法都将剔除模型中重要的变量, 从而导致模型出现偏差;

(3) 然而, 使用替代变量估计 ξ 这一方法和基准统计量的估计方法都能比较准确地选出正确的变量, 从而达到了预期的变量选择效果.

7.5.5 实际应用

本节将上述提出的估计和变量选择方法应用于糖尿病数据分析. 这组数据包含了 442 个数据集, 其中响应变量 Y 是一年之后患者进展情况的测量值. 本例将体重指标 BMI 作为辅助变量 V, 将血清测量值中有关总胆固醇量, 密度水平, 眼压性青光眼水平和血糖浓度作为四个辅助变量 η. 将性别 (SEX) 作为 W, 最后将这三个变量: 年龄 (AGE), 体重 (BMI) 和平均血压 (BP) 作为单指标部分 Z.

直接使用 SCAD-BIC 选择该组数据中的重要变量后, 得到的估计为 $\hat{\alpha}_{\lambda_1} = (-3.12,\ 0, -3.89, 0, 198.45, 0)^{\mathrm{T}}$, $\hat{\beta}_{\lambda_2} = 0$ 和 $\hat{\theta}_{\lambda_3} = (0.88, 0.03, 0.48)^{\mathrm{T}}$. 表明 $\xi_2(V)$, $\xi_4(V)$, $\xi_6(V)$ 和 W 都是不显著的变量, 可从模型中排除.

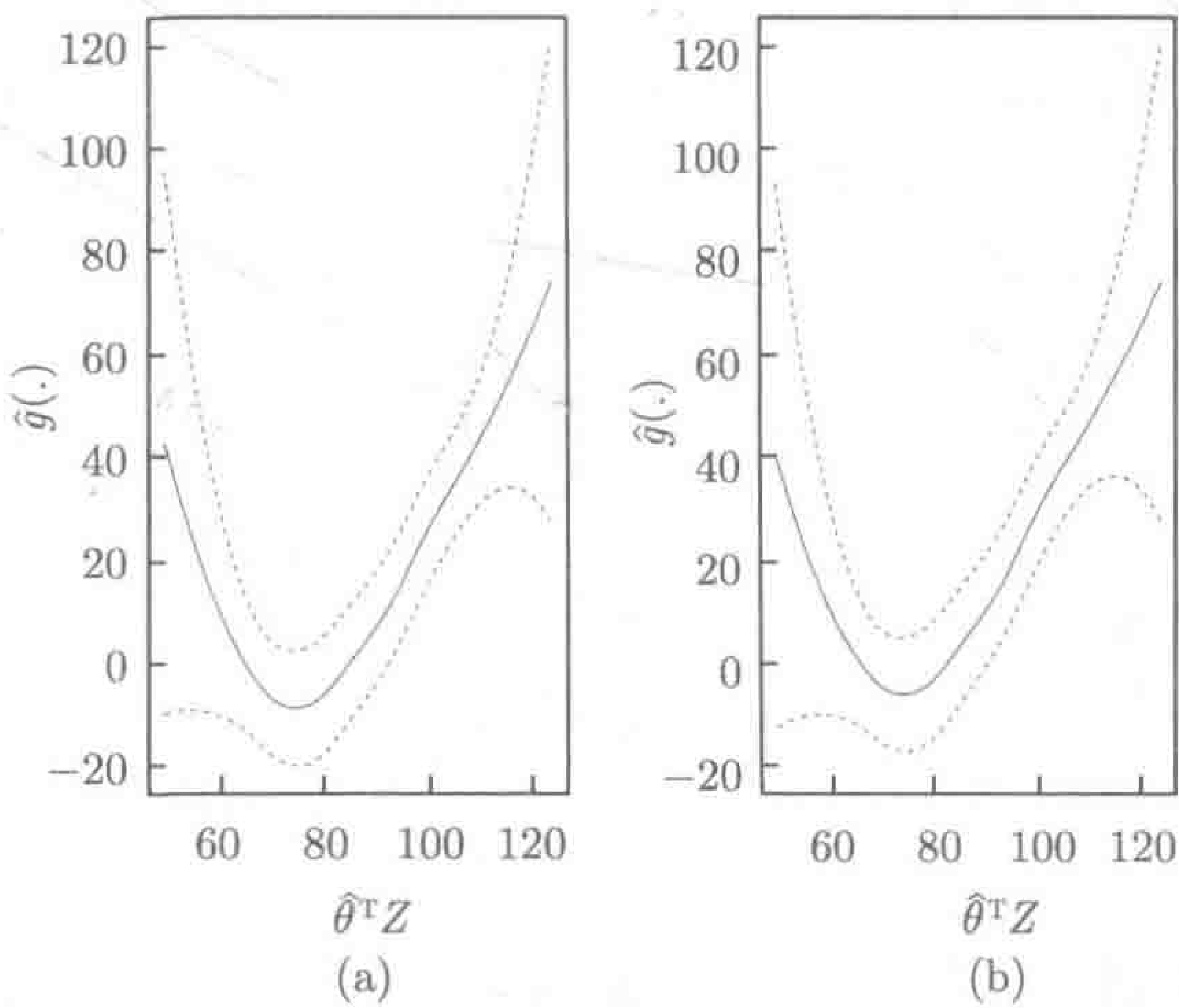


图 7.5.1 实线表示联系函数 $g(\cdot)$ 的局部线性拟合曲线, 虚线表示联系函数 95%的置信带. 横坐标表示 $\hat{\theta}^{\mathrm{T}}Z$, (a) 图是基于惩罚 profile 最小二乘 SCAD-BIC 方法得到曲线, (b) 图是基于 profile 最小二乘方法得到的曲线

基于数据集 $\{(Y_i-\hat{\alpha}_{\lambda_1}^{\mathrm{T}}\hat{\xi}_i-\hat{\beta}_{\lambda_2}^{\mathrm{T}}W_i,\hat{\theta}_{\lambda_3}^{\mathrm{T}}Z_i),i=1,\cdots,n\}$, 继续使用 Fan 和 Gijbels (1996) 提出的局部线性光滑方法来得到联系函数 $g(\cdot)$ 的估计. 图 7.5.1(a) 展示了 $\hat{g}(\cdot)$ 的拟合曲线和 95%的置信带. 此外, 还对比了 profile 最小二乘方法得到的参数估计结果, 其估计结果为 $\hat{\alpha}=(-7.07,5.06,-8.67,-143.04,386.27,1.26)^{\mathrm{T}}$, $\hat{\beta}=-6.83$ 和 $\hat{\theta}=(0.88,0.03,0.48)^{\mathrm{T}}$. 同样地, 可基于数据 $\{(Y_i-\hat{\alpha}^{\mathrm{T}}\hat{\xi}_i-\hat{\beta}^{\mathrm{T}}W_i,\hat{\theta}^{\mathrm{T}}Z_i),i=1,\cdots,n\}$ 来估计联系函数 $g(\cdot)$. 图 7.5.1(b) 展示了 $\hat{g}(\cdot)$ 的拟合曲线和 95%的置信带. 可以看出, 这两个图没有太大的区别, 但是 SCAD-BIC 可去掉模型 (7.5.2) 中不重要的变量, 从而精简了模型.

第 8 章 测量误差降维模型

本章首先介绍几种经典的充分降维方法, 如切片逆回归方法、核逆回归方法、逆方差方法、等高线回归方法、方向回归方法、Hessian 主方向方法、最小平均方差估计方法、条件密度函数的 dMAVE 方法和累积分布降维方法, 并讨论结构维数的选取问题. 然后针对可加测量误差降维模型和协变量调整降维模型讨论充分降维方法及其相关理论.

8.1 引 言

随着现代科学技术和计算机的迅速发展以及统计学在各个科学技术领域中的应用, 很容易收集到维数很高的海量数据, 这种高维数据的统计分析给统计学者带来了极大的困难, 因此高维数据的统计建模和适应性统计方法研究就成为目前统计学研究的具有挑战性的课题. 例如, 人们在分析实际数据时经常会面临协变量 X 的维数 p 很大的情形, 此时如果使用传统的非参数光滑方法建立响应变量 Y 和 p 维协变量 X 之间的回归关系, "维数祸根" 问题将使得传统的非参数光滑估计的效率大大降低. 因此, 如何有效地处理实际数据分析中的高维变量显得尤为重要. 充分降维 (sufficient dimension reduction, SDR) 方法就是一种非常行之有效的方法. 具体来说, 就是在不丢失太多信息的前提下, 把 p 个协变量进行线性组合, 然后用组合后的 $k(k \leqslant p)$ 个线性组合来替代原来的 p 维协变量进行统计分析.

充分降维方法 (SDR) 的目的在于寻找一个矩阵 B, 使得 响应变量 Y 和协变量 X 满足独立性条件:

$$Y \perp\!\!\!\perp X | B^{\mathrm{T}} X, \tag{8.1.1}$$

其中 $\perp\!\!\!\perp$ 表示独立, $B = (\beta_1, \cdots, \beta_k)$ 是一个 $p \times k$ 维的列满秩常数矩阵, 这里 β_s 是 p 维非 0 列向量, $s = 1, \cdots, k$. 等价地, 模型 (8.1.1) 也可以写成如下形式:

$$Y = G(B^{\mathrm{T}} X, \varepsilon), \tag{8.1.2}$$

其中 ε 是模型误差, 与协变量 X 相互独立, $G(\cdot,\cdot)$ 是未知的联系函数. 换句话说, 给定 $B^{\mathrm{T}} X$ 条件下, Y 和 X 相互独立. 此外, 矩阵 B 不是唯一的, 因为对矩阵 B 进行任意的正交变换仍能满足条件独立性 (8.1.1). 因此, 充分降维方法关心的是矩阵 B 的列向量所张成的线性空间 $\mathcal{S}(B)$ 而不是矩阵 B 本身, 通常也把线性空间 $\mathcal{S}(B)$

称为降维空间, 并且把列满秩的 $p\times k$ 维常数矩阵 B 中的 k 称为降维空间 $\mathcal{S}(B)$ 的结构维数.

不失一般性, 可假定降维矩阵 B 满足条件 $B^{\mathrm{T}}B=I_k$, 其中 I_k 是 $k\times k$ 维单位矩阵, 即降维矩阵 B 的列向量相互正交, $\|\beta_s\|=1$, $s=1,\cdots,k$, $\beta_s^{\mathrm{T}}\beta_t=0$, $1\leqslant s\neq t\leqslant k$.

本章内容结构如下.

8.2 节介绍一些常用的充分降维方法和结构维数 k 的选取问题, 重点介绍切片逆回归方法、核逆回归方法、逆方差方法、等高线回归方法、方向回归方法、Hessian 主方向方法、最小平均方差估计方法、条件密度函数的 dMAVE 方法和累积分布降维方法.

8.3 节考察测量误差结构, 协变量 X 不能被准确观测到, 而可观测到的是变量 W, 并且观测变量 W 与不可观测的协变量 X 有如下的关系式:

$$W=\gamma+\Gamma X+\delta, \tag{8.1.3}$$

其中 $X=(X_1,\cdots,X_p)^{\mathrm{T}}$ 是 $p\times 1$ 的不可观测协变量, γ 是 $r\times 1$ 的常数向量, Γ 是 $r\times p$ 的常数矩阵, 并且有 $r\geqslant p$, δ 是一个 r 维的误差向量, 并假设 W 和 X 相互独立. 8.3 节主要关心的是如何利用观测到的 (Y,W) 来求解降维模型 (8.1.2) 中由矩阵 B 的列向量所张成的降维空间 $\mathcal{S}(B)$ 的估计问题.

8.4 节考虑如下的协变量调整降维模型:

$$\begin{cases} Y \perp\!\!\!\perp X|B^{\mathrm{T}}X, \\ \widetilde{Y}=\phi(U)Y, \\ \widetilde{X}_r=\psi_r(U)X_r, \quad r=1,\cdots,p, \end{cases} \tag{8.1.4}$$

其中 $\perp\!\!\!\perp$ 表示独立, Y 是不可观测的响应变量, $X=(X_1,X_2,\cdots,X_p)^{\mathrm{T}}$ 是不可观测的连续型协变量向量, $\widetilde{Y}$ 和 $\widetilde{X}_r$ 分别是观测到的响应变量和协变量, U 是一个 1 维可观测的混淆变量, $\phi(U)$ 和 $\psi_r(U)$ 是关于混淆变量 U 的未知失真函数. 8.4 节主要关心的是如何利用观测到的变量 $(\widetilde{Y},\widetilde{X},U)$ 求解降维模型 (8.1.4) 中由矩阵 B 的列向量所张成的降维空间 $\mathcal{S}(B)$ 的估计问题.

8.2 充分降维方法

在介绍测量误差降维模型估计问题之前, 本节系统介绍常用的一些估计降维空间 $\mathcal{S}(B)$ 的充分降维方法, 如 Li (1991) 提出的切片逆回归 (sliced inverse regression, SIR) 方法, Zhu 和 Fang (1996) 提出的核逆回归 (kernel inverse regression, KIR) 方法, Cook (1998a) 和 Cook (2000) 提出的切片方差回归方法, Li 等 (2005) 提出的等

高线回归方法, Li 和 Wang (2007) 提出的直接回归方法, Li (1992) 提出的 Hessian 主方向 (principal Hessian direction, PHD) 降维方法, Xia 等 (2002) 提出的最小平均方差估计 (MAVE), Xia (2007) 提出的基于条件密度函数的最小平均方差估计方法, Zhu 等 (2010) 提出的累积切片估计方法. 这些估计方法都能得到降维空间 $\mathcal{S}(B)$ 的 $\sqrt{n}$ 相合估计并且计算速度也很快. 值得一提的是, 在实际数据分析中, 一旦得到了 B 的估计 $\hat{B}$, 便可使用现有的统计方法来分析较低维数的变量 $\hat{B}^{\mathrm{T}}X$, 如使用变量 $(Y,\hat{B}^{\mathrm{T}}X)$ 建立适当的模型 (如线性模型、非线性模型等) 来分析数据.

8.2.1 逆回归方法

逆回归方法 (inverse regression, IR) 的核心思想是通过计算 $E(X|Y)$ 来估计降维空间 $\mathcal{S}(B)$, 其中 Y 是 1 维响应变量, X 是 p 维协变量. 不妨假设 $E(Y|X)=g(B^{\mathrm{T}}X)$, $g(\cdot)$ 是未知的联系函数, B 是未知的降维矩阵. 当在协变量 X 的维数过高时, 若直接建立变量 (Y,X) 之间的非参数核光滑方法 (Fan and Gijbels, 1996) 或者样条光滑方法 (Hastie and Tibshirani, 1990; Stone, 1985, 1986; Li and Ruppert, 2008; Zhang and Liang, 2011) 来给出降维矩阵 B 的估计, 此时会遇到传统非参数回归的 “维数祸根” 问题. 逆回归方法提供了一种解决方法, 该方法的一个最重要的特点就是通过计算 $E(X|Y)$ 来给出降维矩阵 B 的估计. 不妨记 $E(X|Y=y)=(g_{X_1}(y),\cdots,g_{X_p}(y))^{\mathrm{T}}$. 在 Y 是 1 维的响应变量时, 此时关于 $g_{X_s}(y)$ 的估计可用传统的 1 维非参数光滑方法来给出, 避免了 “维数祸根” 问题的出现.

在介绍逆回归方法之前, 不妨假设 $E(X)=0$, 并引入关于协变量 X 自身的一个线性条件:

$$E\left(X|B^{\mathrm{T}}X\right)=P_B^{\mathrm{T}}(\Sigma_X)X,\quad P_B(\Sigma_X)=B\left(B^{\mathrm{T}}\Sigma_XB\right)^{-1}B^{\mathrm{T}}\Sigma_X, \tag{8.2.1}$$

其中 $\Sigma_X=\mathrm{Var}(X)$ 是协变量 X 的协方差矩阵. 在线性条件 (8.2.1) 下, 对任意向量 $b\in\mathbb{R}^p$, 有

$$\begin{aligned}E(b^{\mathrm{T}}X|B^{\mathrm{T}}X)&=E(b^{\mathrm{T}}X|\beta_1^{\mathrm{T}}X,\beta_2^{\mathrm{T}}X,\cdots,\beta_k^{\mathrm{T}}X)\\&=c_1\beta_1^{\mathrm{T}}X+c_2\beta_2^{\mathrm{T}}X+\cdots+c_k\beta_k^{\mathrm{T}}X,\end{aligned} \tag{8.2.2}$$

其中 $c_1,\cdots,c_k$ 是常数并满足 $(c_1,\cdots,c_k)^{\mathrm{T}}=(B^{\mathrm{T}}\Sigma_XB)^{-1}B^{\mathrm{T}}\Sigma_Xb$. 在总体意义下, 条件数学期望或者说回归函数 $E(b^{\mathrm{T}}X|B^{\mathrm{T}}X)$ 是 $B^{\mathrm{T}}X$ 的线性函数. 线性条件不是很强的条件, 若协变量 X 服从椭球分布 (Eaton, 1986), 如正态分布或者 t 分布, 线性条件 (8.2.1) 自然成立. Hall 和 Li (1993) 证明了, 当协变量 X 的维数 p 趋于无穷时, 线性条件 (8.2.1) 可近似成立. 因此, 在高维数据背景下, 线性条件完全可看成是一个基本假设条件.

在线性条件 (8.2.1) 和独立性条件 (8.1.1) 成立的情况下, Li (1991) 给出了如下事实:

$$
\begin{aligned}
E(X|Y) &= E\left[E(X|Y)\Big|\left(B^{\mathrm{T}}X,Y\right)\right] = E\left[E\left[X|\left(B^{\mathrm{T}}X,Y\right)\right]\Big|Y\right] \\
&= E\left[E\left[X|B^{\mathrm{T}}X\right]\Big|Y\right] = E\left[P_B^{\mathrm{T}}(\Sigma_X)X\Big|Y\right] \\
&= \Sigma_X B\left(B^{\mathrm{T}}\Sigma_X B\right)^{-1}B^{\mathrm{T}}E(X|Y).
\end{aligned} \tag{8.2.3}
$$

式 (8.2.3) 表明 $\Sigma_X^{-1}E(X|Y) = B\left(B^{\mathrm{T}}\Sigma_X B\right)^{-1}B^{\mathrm{T}}E(X|Y) \subseteq \mathcal{S}(B)$, 即 $\Sigma_X^{-1}E(X|Y)$ 是矩阵 B 列向量 $\beta_1,\cdots,\beta_k$ 的线性组合. 不难看出, 当结构维数 $k=1$ 时, 此时有 $B=\beta_1$, 式 (8.2.3) 可简化为

$$
E(X|Y) = \frac{E(\beta_1^{\mathrm{T}}X|Y)}{\beta_1^{\mathrm{T}}\Sigma_X\beta_1}\Sigma_X\beta_1. \tag{8.2.4}
$$

若记 $c(Y) = \dfrac{E(\beta_1^{\mathrm{T}}X|Y)}{\beta_1^{\mathrm{T}}\Sigma_X\beta_1}$, 式 (8.2.4) 表明 $\Sigma_X^{-1}E(X|Y)$ 与 β_1 成正比例关系, 即

$$
\Sigma_X^{-1}E(X|Y) = c(Y)\beta_1.
$$

定义逆回归降维矩阵 Λ_{IR} 如下:

$$
\Lambda_{\mathrm{IR}} =: \Sigma_X^{-1}E_Y\left[E(X|Y)E(X^{\mathrm{T}}|Y)\right]\Sigma_X^{-1}, \tag{8.2.5}
$$

其中 $E_Y(\cdot)$ 表示对 Y 求数学期望.

在结构维数 $k=1$ 时, $\Lambda_{\mathrm{IR}} = E[c^2(Y)]\beta_1\beta_1^{\mathrm{T}}$. 若 $E[c^2(Y)] \neq 0$, 从式 (8.2.5) 不难看出逆回归降维矩阵 Λ_{IR} 的秩为 1. 若进一步对 Λ_{IR} 进行特征值分解, $\Lambda_{\mathrm{IR}}\eta_{\mathrm{IR}} = \lambda_{\max}\eta_{\mathrm{IR}}$, 其中 $\lambda_{\max}$ 是最大非零特征值, η_{IR} 是最大非零特征值 $\lambda_{\max}$ 所对应的特征向量. 再结合条件 $\|\beta_1\|=1$ 和特征向量模长为 1, 可以证明 $\beta_1^{\mathrm{T}}\eta = \pm 1$, 进而有

$$
\beta_1 = \pm\eta_{\mathrm{IR}},
$$

即降维矩阵 Λ_{IR} 的最大特征向量 η_{IR} 和 β_1 至多差一个符号. 当结构维数 $k=1$ 时, 降维矩阵 Λ_{IR} 的最大特征向量 η_{IR} 可以用于估计 β_1 所张成的线性空间, 即 $\mathcal{S}(\eta_{\mathrm{IR}}) = \mathcal{S}(\beta_1)$.

当结构维数 $k>1$ 时, 并且结构维数 k 已知时, 可对降维矩阵 Λ_{IR} 进行如下特征值分解:

$$
\Lambda_{\mathrm{IR}}\eta_s = \lambda_s\eta_s, \quad \lambda_1 \geqslant \lambda_2 \geqslant \cdots \geqslant \lambda_k > 0, \tag{8.2.6}
$$

其中 η_s 是第 s 个非零特征值 λ_s 所对应的特征向量. Li (1991) 证明了在线性条件 (8.2.3) 成立的情况下, $\mathcal{S}(\eta_1,\cdots,\eta_k)\subseteq\mathcal{S}(B)$, 即由前 k 个特征向量 $\eta_1,\cdots,\eta_k$ 所张成的列向量空间可用于求解降维空间 $\mathcal{S}(B)$.

下面将介绍两种逆回归估计方法, 分别是 Li (1991) 提出的切片逆回归方法与 Zhu 和 Fang (1996) 提出的核逆回归方法.

1. *切片逆回归方法*

假设 $\{(X_i,Y_i),i=1,\cdots,n\}$ 是来自于模型 (8.1.1) 的一组 i.i.d.的随机样本. Li (1991) 提出切片逆回归的方法来估计降维矩阵 Λ_{IR}. 具体算法如下.

步骤 1 先将响应变量的样本数据 $Y_1,\cdots,Y_n$ 按照其取值从小到大排序 $Y_{(1)}\leqslant Y_{(2)}\leqslant\cdots\leqslant Y_{(n)}$, 并把排序后的数据等分为 H 个区间 (也称为 H 个切片), 并将区间分别记为 $B_{n,1},\cdots,B_{n,H}$, 如

$$B_{n,1}=\{Y_{(1)},Y_{(2)},\cdots,Y_{(m)}\},\quad B_{n,2}=\{Y_{(m+1)},Y_{(m+2)},\cdots,Y_{(2m)}\},\quad\cdots,$$

依次类推. 通常情况下, 区间中的数据个数 m 的取值范围应满足

$$m\in\left(\min\left\{\sqrt{n},\frac{n}{2}\right\},\max\left\{\sqrt{n},\frac{n}{2}\right\}\right).$$

若最后一个区间 $B_{n,H}$ 包含的数据量过少 (如只有一个或者两个数据), 这时可将区间 $B_{n,H}$ 中的数据合并到区间 $B_{n,H-1}$ 中. Hsing 和 Carroll (1992), Zhu 和 Ng (1995) 研究了切片数 H 的选取和相关的理论性质, 有兴趣的读者可参考相关文献.

步骤 2 将排序后的数据 $Y_{(m)}$ 和原始数据 (X_s,Y_s) 进行匹配, 其中 $Y_s=Y_{(m)}$, 并将此时的数据 X_s 记为 $X_{(:,m)}$, 即 $(X_{(:,m)},Y_{(m)})=(X_s,Y_s)$.

步骤 3 计算每个区间 (或切片)$B_{n,l}$ 内与响应变量相对应的协变量 X 的均值, 即

$$\bar{X}^{(l)}=\frac{1}{n_l}\sum_{i:Y_{(i)}\in B_{n,l}}X_{(:,i)},$$

其中 n_l 是第 l 个区间 (或切片) $B_{n,l}$ 内包含的样本观测个数, $l=1,\cdots,H$.

步骤 4 给出降维矩阵 Λ_{IR} 的切片逆回归估计 $\hat{\Lambda}_{\mathrm{SIR}}$ 和协方差阵 Σ_X 的估计 $\hat{\Sigma}_X$, 分别定义如下

$$\hat{\Lambda}_{\mathrm{SIR}}=\hat{\Sigma}_X^{-1}\left(\frac{1}{n}\sum_{l=1}^{H}n_l\left(\bar{X}^{(l)}-\bar{X}\right)\left(\bar{X}^{(l)}-\bar{X}\right)^{\mathrm{T}}\right)\hat{\Sigma}_X^{-1},$$

$$\hat{\Sigma}_X=\frac{1}{n}\sum_{i=1}^{n}\left(X_i-\bar{X}\right)\left(X_i-\bar{X}\right)^{\mathrm{T}},\quad\bar{X}=\frac{1}{n}\sum_{i=1}^{n}X_i.$$

步骤 5　计算切片逆回归方向向量 $\hat{\eta}_1, \cdots, \hat{\eta}_k$:

$$\hat{\Lambda}_{\mathrm{SIR}}\hat{\eta}_s = \hat{\lambda}_s\hat{\eta}_s, \quad \hat{\lambda}_1 \geqslant \hat{\lambda}_2 \geqslant \cdots \geqslant \hat{\lambda}_k > 0.$$

这时, 也称 $\hat{\eta}_s$ 为切片逆回归的第 s 个方向向量.

切片逆回归方法先将数据 $\{Y_1, \cdots, Y_n\}$ 进行排序后, 并将排序后的数据分成 H 个切片, 其目的是为了计算条件数学期望 $E(X|Y)$. Li (1991) 证明了在一定的正则条件下, 切片逆回归方向向量所张成的空间 $\mathcal{S}(\hat{\eta}_1, \cdots, \hat{\eta}_k)$ 是降维空间 $\mathcal{S}(B)$ 的一个 $\sqrt{n}$ 相合估计.

2. *核逆回归方法*

Zhu 和 Fang (1996) 提出了核逆回归方法, 主要思想是可以采用经典的非参数核光滑估计方法对条件数学期望 $E(X|Y=\cdot)$ 进行估计, 进而给出降维矩阵 Λ_{IR} 的核估计. 假设 $\{(X_i, Y_i), i = 1, \cdots, n\}$ 是来自于模型 (8.1.1) 的一组 i.i.d.的随机样本. 具体估计方法如下.

步骤 1　使用局部常数 Nadaraya-Watson 核光滑方法给出 $g(y) =: E(X|Y=y)$ 的估计, 定义为

$$\hat{g}(y) = \frac{\sum_{i=1}^{n} K_h(Y_i - y)X_i}{\sum_{i=1}^{n} K_h(Y_i - y)},$$

其中 $K_h(\cdot) = h^{-1}K(\cdot/h)$, $K(\cdot)$ 是核密度函数, h 是窗宽.

步骤 2　给出降维矩阵 Λ_{IR} 的核估计 $\hat{\Lambda}_{\mathrm{KIR}}$ 和协方差阵 Σ_X 的估计 $\hat{\Sigma}_X$, 分别定义如下:

$$\hat{\Lambda}_{\mathrm{KIR}} = \hat{\Sigma}_X^{-1}\left(\frac{1}{n}\sum_{i=1}^{n}\hat{g}(Y_i)\hat{g}^{\mathrm{T}}(Y_i)\right)\hat{\Sigma}_X^{-1},$$

$$\hat{\Sigma}_X = \frac{1}{n}\sum_{i=1}^{n}\left(X_i - \bar{X}\right)\left(X_i - \bar{X}\right)^{\mathrm{T}}, \quad \bar{X} = \frac{1}{n}\sum_{i=1}^{n}X_i.$$

步骤 3　计算核逆回归方向向量 $\hat{\mu}_1, \cdots, \hat{\mu}_k$ 如下:

$$\hat{\Lambda}_{\mathrm{KIR}}\hat{\mu}_s = \hat{\lambda}_s\hat{\mu}_s, \quad \hat{\lambda}_1 \geqslant \hat{\lambda}_2 \geqslant \cdots \geqslant \hat{\lambda}_k > 0.$$

这时, 也称 $\hat{\mu}_s$ 为核逆回归的第 s 个方向向量, $s = 1, \cdots, k$.

Zhu 和 Fang (1996) 证明了在一定的正则条件下, 核逆回归方向向量所张成的空间 $\mathcal{S}(\hat{\mu}_1, \cdots, \hat{\mu}_k)$ 是降维空间 $\mathcal{S}(B)$ 的一个 $\sqrt{n}$ 相合估计.

除了切片逆回归方法和核逆回归方法之外, Hsing (1999) 使用 K 近邻非参数方法提出了近邻逆回归方法. Fung 等 (2002) 采用 B 样条方法对 $E(X|Y=\cdot)$ 进行估计, 并提出了基于典则相关系数的 SIR 降维方法. 此外, SIR 方法在其他统计分支中也受到了很大的关注. Carroll 和 Li (1992) 提出了基于 SIR 方法的测量误差降维模型, Li 等 (1999) 将 SIR 方法应用于响应变量 Y 有删失的情形, Li 等 (2003) 提出如何把 SIR 方法推广到响应变量 Y 是多维的情形, Zhu 等 (2006) 将 SIR 方法拓展到协变量 X 的维数是发散情形下的高维数据分析中.

8.2.2 逆方差方法

Cook (1998a) 注意到, 在一些特殊情况下, 逆回归方法有时并不能有效地找到降维空间 $\mathcal{S}(B)$. Cook (1998a) 给出下面的一个例子. 不妨假设结构维数 $k=1$, 协变量 X 服从 p 维标准正态分布, 并且假定在给定 X 的情况下, $Y|X$ 的条件分布服从正态分布 $N\left((\beta_1^{\mathrm{T}}X)^2,\sigma^2\right)$, 其中 β_1 是非 0 向量, $\sigma>0$. 等价地, 可以表示为

$$Y|X=(\beta_1^{\mathrm{T}}X)^2+\sigma\varepsilon,\quad X\text{与}\varepsilon\text{相互独立},\quad \varepsilon\sim N(0,1).$$

如图 8.2.1 所示, 由对称性可知条件数学期望 $E(\beta_1^{\mathrm{T}}X|Y)=0$, 再根据式 (8.2.4), 可以得到 $E(X|Y)=0$ 和 $\Lambda_{\mathrm{IR}}=0$. 即 $\Lambda_{\mathrm{IR}}=0$ 不存在非零特征值, 也就不可能通过逆回归的方法来估计 β_1 所张成的降维空间 $\mathcal{S}(\beta_1)$.

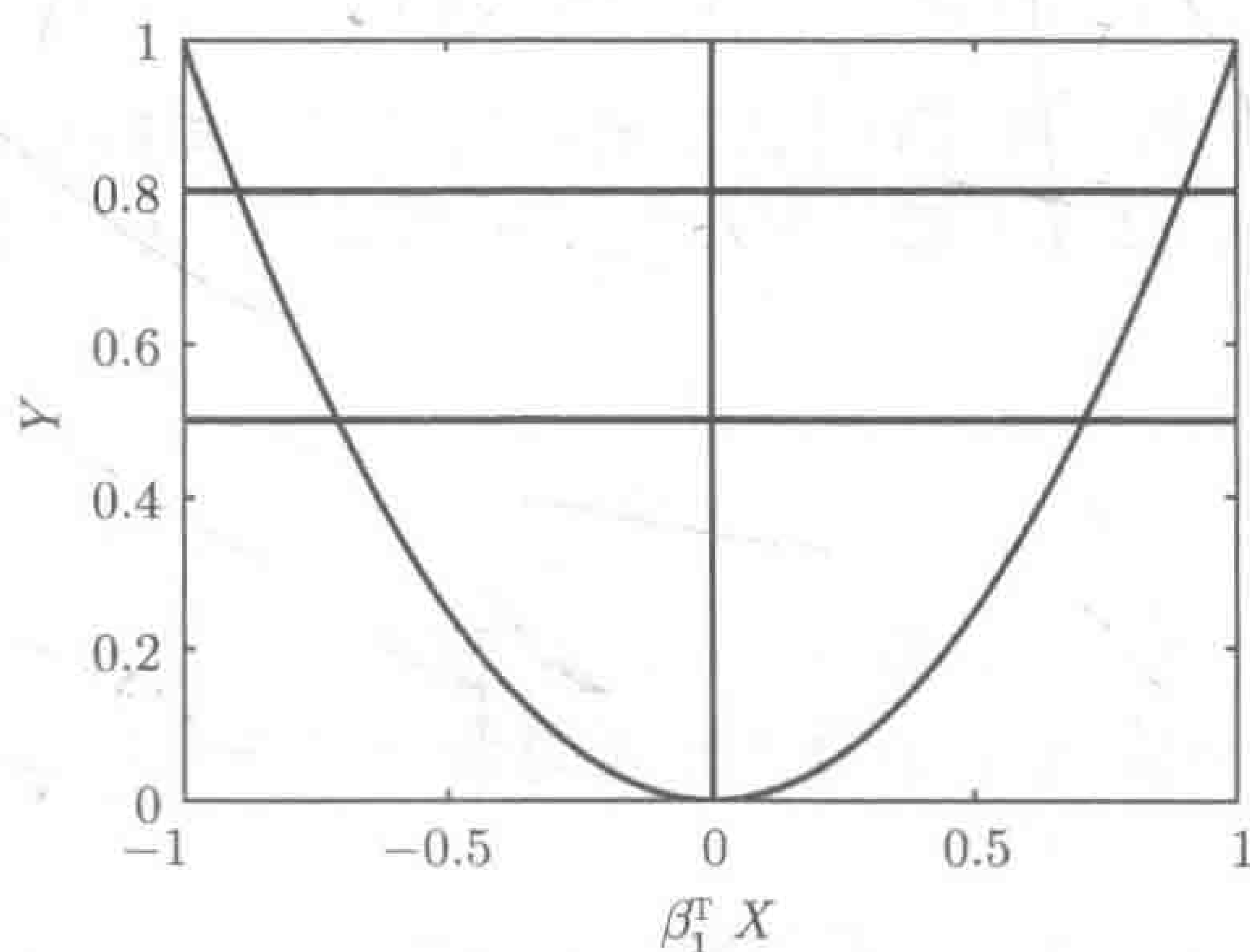


图 8.2.1 抛物线表示 $E(Y|X)=(\beta_1^{\mathrm{T}}X)^2$, 两条横线分别表示 $Y=0.5$ 和 $Y=0.8$

虽然不可能通过式 (8.2.4) 中的条件期望 $E(X|Y)$ 来得到有关 β_1 的信息, 但是结合图 8.2.1, Cook (1998a) 注意到条件方差 $\mathrm{Cov}(\beta_1^{\mathrm{T}}X|Y=y)$ 会随着 y 的变化而变化. 因此 Cook (1998a) 提出基于条件方差矩阵 $\mathrm{Cov}(X|Y)$ 来建立矩阵 $\mathrm{Cov}(X|Y)$ 和 β_1 之间的关系. 因此, Cook (1998a) 引入如下的常数方差条件:

$$\mathrm{Cov}(X|B^{\mathrm{T}}X)=\Sigma_X-\Sigma_X B\left(B^{\mathrm{T}}\Sigma_X B\right)^{-1}B^{\mathrm{T}}\Sigma_X=\Sigma_X-P_B^{\mathrm{T}}(\Sigma_X)\Sigma_X. \tag{8.2.7}$$

当协变量 X 服从正态分布时, 常数方差条件成立 (Eaton, 1986). 当协变量 X 服从其他的一些椭球等高分布时, 常数方差条件也近似地成立 (Cook, 1998a).

不妨假设 $EX=0$, 在线性条件 (8.2.1) 和常数方差条件 (8.2.7) 成立的情况下, Cook (1998a, 2000) 得到如下关系式:

$$\Sigma_X-\mathrm{Cov}(X|Y)=P_B^{\mathrm{T}}(\Sigma_X)\Sigma_X+P_B^{\mathrm{T}}(\Sigma_X)\mathrm{Cov}(X|Y)P_B(\Sigma_X). \tag{8.2.8}$$

不难看出 $\Sigma_X^{-1}\left[\Sigma_X-\mathrm{Cov}(X|Y)\right]\subseteq\mathcal{S}(B)$. 特殊地, 当结构维数 $k=1$, 即 $B=\beta_1$ 时,

$$\begin{aligned}\Sigma_X-\mathrm{Cov}(X|Y)&=\left[1+\beta_1^{\mathrm{T}}\mathrm{Cov}(X|Y)\beta_1\right]\Sigma_X\beta_1\beta_1^{\mathrm{T}}\Sigma_X\\&=:d(Y)\Sigma_X\beta_1\beta_1^{\mathrm{T}}\Sigma_X,\end{aligned} \tag{8.2.9}$$

其中 $d(Y)$ 是 1 维的变量.

基于式 (8.2.8), Cook (1998a, 2000) 定义了如下的降维矩阵 Λ_{SA}, 即

$$\Lambda_{\mathrm{SA}}=:\Sigma_X^{-1}E_Y\left[\{\Sigma_X-\mathrm{Cov}(X|Y)\}\{\Sigma_X-\mathrm{Cov}(X|Y)\}^{\mathrm{T}}\right]\Sigma_X^{-1},$$

其中 $E_Y(\cdot)$ 表示对 Y 取数学期望.

进一步当结构维数 $k=1$, 即 $B=\beta_1$ 时, 基于式 (8.2.9) 可得到 $\Lambda_{\mathrm{SA}}=c_0\beta_1\beta_1^{\mathrm{T}}$, 常数 $c_0=E[d^2(Y)]\beta_1^{\mathrm{T}}\Sigma_X\Sigma_X\beta_1$. 不难看出, 在 $c_0\neq 0$ 成立的情况下, 降维矩阵 Λ_{SA} 是秩为 1 的对称矩阵. 若进一步对降维矩阵 Λ_{SA} 进行特征值分解, $\Lambda_{\mathrm{SA}}\eta_{\mathrm{SA}}=\lambda_{\max}\eta_{\mathrm{SA}}$, 其中 $\lambda_{\max}$ 是最大非零特征值, η_{SA} 是最大非零特征值 $\lambda_{\max}$ 所对应的特征向量. 再结合条件 $\|\beta_1\|=1$ 和特征向量模长为 1, 可以证明 $\beta_1^{\mathrm{T}}\eta=\pm 1$. 那么有下面关系式成立:

$$\beta_1=\pm\eta_{\mathrm{SA}},$$

即降维矩阵 Λ_{SA} 的最大特征向量 η_{SA} 和 β_1 至多差一个符号. 此时, 在结构维数 $k=1$ 的情形下, 降维矩阵 Λ_{SA} 的最大特征向量 η_{SA} 可以用于求解 β_1 所张成的线性空间, 即 $\mathcal{S}(\eta_{\mathrm{SA}})=\mathcal{S}(\beta_1)$ 成立.

当结构维数 $k>1$ 时, 并且结构维数 k 已知时, 可对降维矩阵 Λ_{SA} 进行如下特征值分解, 即

$$\Lambda_{\mathrm{SA}}\eta_s=\lambda_s\eta_s,\quad \lambda_1\geqslant\lambda_2\geqslant\cdots\geqslant\lambda_k>0,$$

其中 η_s 是第 s 个非零特征值 λ_s 所对应的特征向量. Cook(1998a, 2000) 证明了在线性条件 (8.2.3) 和常数方差条件 (8.2.7) 成立的情况下, $\mathcal{S}(\eta_1,\cdots,\eta_k)\subseteq\mathcal{S}(B)$, 即由前 k 个特征向量 $\eta_1,\cdots,\eta_k$ 所张成的列向量空间可用于求解降维空间 $\mathcal{S}(B)$.

假设 $\{(X_i,Y_i),i=1,\cdots,n\}$是来自于模型 (8.1.1) 的一组 i.i.d.的随机样本.

Cook (2000) 提出了切片平均方差估计方法(sliced average inverse variance estimation, SAVE) 来估计降维矩阵 Λ_{SA}, 具体算法如下.

步骤 1 先将响应变量的样本数据 $Y_1,\cdots,Y_n$ 按照其取值从小到大排序, 即 $Y_{(1)}\leqslant Y_{(2)}\leqslant\cdots\leqslant Y_{(n)}$, 然后把排序后的数据等分为 H 个区间 (也称为 H 个切片), 分别将区间记为 $B_{n,1},\cdots,B_{n,H}$, 如

$$B_{n,1}=\{Y_{(1)},\cdots,Y_{(m)}\},\quad B_{n,2}=\{Y_{(m+1)},Y_{(m+2)},\cdots,Y_{(2m)}\},\quad\cdots,$$

依次类推.

步骤 2 将排序后的数据 $Y_{(m)}$ 和原始数据 (X_s,Y_s) 进行匹配, 其中 $Y_s=Y_{(m)}$, 并将此时的数据 X_s 记为 $X_{(:,m)}$, 即 $(X_{(:,m)},Y_{(m)})=(X_s,Y_s)$.

步骤 3 计算每个区间 (或切片)$B_{n,l}$ 内与响应变量相对应的协变量 X 的协方差矩阵, 定义为

$$\widehat{\mathrm{Cov}}(X|Y\in B_{n,l})=\frac{1}{n_l}\sum_{i:Y_{(i)}\in B_{n,l}}\left(X_{(:,i)}-\bar{X}^{(l)}\right)\left(X_{(:,i)}-\bar{X}^{(l)}\right)^{\mathrm{T}},$$

其中 $\bar{X}^{(l)}=\dfrac{1}{n_l}\displaystyle\sum_{i:Y_{(i)}\in B_{n,l}}X_{(:,i)}$, n_l 是第 l 个区间 (或切片) $B_{n,l}$ 内的样本观测个数, $l=1,\cdots,H$.

步骤 4 给出降维矩阵 Λ_{SA} 的估计 $\hat{\Lambda}_{\mathrm{SA}}$ 和协方差阵 Σ_X 的估计 $\hat{\Sigma}_X$, 分别定义如下

$$\hat{\Lambda}_{\mathrm{SA}}=\hat{\Sigma}_X^{-1}\left(\frac{1}{n}\sum_{l=1}^{H}n_l\left(\hat{\Sigma}_X-\widehat{\mathrm{Cov}}(X|Y\in B_{n,l})\right)\left(\hat{\Sigma}_X-\widehat{\mathrm{Cov}}(X|Y\in B_{n,l})\right)^{\mathrm{T}}\right)\hat{\Sigma}_X^{-1},$$

$$\hat{\Sigma}_X=\frac{1}{n}\sum_{i=1}^{n}\left(X_i-\bar{X}\right)\left(X_i-\bar{X}\right)^{\mathrm{T}},\quad \bar{X}=\frac{1}{n}\sum_{i=1}^{n}X_i.$$

步骤 5 计算切片逆回归方向向量 $\hat{\eta}_1,\cdots,\hat{\eta}_k$, 即

$$\hat{\Lambda}_{\mathrm{SA}}\hat{\eta}_s=\hat{\lambda}_s\hat{\eta}_s,\quad \hat{\lambda}_1\geqslant\hat{\lambda}_2\geqslant\cdots\geqslant\hat{\lambda}_k>0.$$

称 $\hat{\eta}_s$ 为 SAVE 的第 s 个方向向量, Cook(1998a, 2000) 详细研究了 SAVE 方法的大样本性质. Li 和 Zhu (2007) 系统研究了 SAVE 方法中切片数选取的相关理论性质. Zhu 和 Zhu (2009a) 利用非参数核光滑方法给出了条件协方差矩阵 $\mathrm{Cov}(X|Y)$ 和降维矩阵 Λ_{SA} 的估计, 并在一定的正则条件下证明了所得估计量的大样本性质. Lue (2008) 提出了适用于删失数据的 SAVE 降维方法. Shao 等 (2007) 提出了关于降维矩阵 Λ_{SA} 结构维数 k 的检验方法, 构造了相应的检验统计量, 并证明了检验统计量的渐近性质.

8.2.3 等高线回归方法

Li 等 (2005) 构造了等高线回归降维矩阵来寻找降维空间 $\mathcal{S}(B)$ 的方法, 该方法的核心思想在于考察当响应变量在给定小范围内变动时, 通过协变量的变动情况来寻找降维空间 $\mathcal{S}(B)$ 的一种估计方法, Li 等 (2005) 把该方法称为等高线回归 (contour regression, CR) 方法.

首先介绍 Li 等 (2005) 引入条件方差的假设条件. 记 $\mathcal{S}^{\perp}(B)$ 是 $\mathcal{S}(B)$ 的正交补空间. 若向量 γ 和 ω 满足: $\forall\gamma\in\mathcal{S}(B)$, $\gamma\in\mathbb{R}^p$ 和 $\forall\omega\in\mathcal{S}^{\perp}(B)$, $\omega\in\mathbb{R}^p$, $\|\gamma\|=\|\omega\|=1$ 时, 存在常数 $c>0$ 使得下面不等式成立:

$$\mathrm{Var}\left(\omega^{\mathrm{T}}(X_1-X_2)\big||Y_1-Y_2|\leqslant c\right)>\mathrm{Var}\left(\gamma^{\mathrm{T}}(X_1-X_2)\big||Y_1-Y_2|\leqslant c\right),\quad(8.2.10)$$

其中 (X_1,Y_1) 和 (X_2,Y_2) 相互独立, 并且和 (X,Y) 有相同的分布函数.

Li 等 (2005) 关于条件方差假设条件 (8.2.10) 列举了大量的数值例子进行说明. 下面以线性模型为例来说明条件方差假设条件 (8.2.10) 的合理性. 考虑线性模型 $Y=\beta_1^{\mathrm{T}}X+\varepsilon$, 其中 X 和 ε 相互独立, $\|\beta_1\|=1$. 若 (X_1,Y_1) 和 (X_2,Y_2) 是来自该模型的两个独立同分布的样本, 那么有

$$Y_1-Y_2=\beta_1(X_1-X_2)+\varepsilon_1-\varepsilon_2.\quad(8.2.11)$$

从模型 (8.2.11) 中不难看出, 当 Y_1-Y_2 的取值在小范围内变化时 (如 $|Y_1-Y_2|\leqslant c$, 其中 c 是一个取值较小的正常数), 那么 $\beta_1(X_1-X_2)$ 的变化也比较小, 即 $\mathrm{Var}(\beta_1^{\mathrm{T}}(X_1-X_2)\big||Y_1-Y_2|\leqslant c)$ 的取值也会较小. 此外, 若 ω 满足 $\omega^{\mathrm{T}}\beta_1=0$, 即 $\omega\in\mathcal{S}^{\perp}(\beta_1)$. 因为 $\omega^{\mathrm{T}}X$ 和模型 (8.2.11) 没有直接关系, 此时当 Y_1-Y_2 的取值在小范围内变化时, 对 $\omega^{\mathrm{T}}(X_1-X_2)$ 的变化不会有太大的影响, 即相对于 $\mathrm{Var}\left(\beta_1^{\mathrm{T}}(X_1-X_2)\big||Y_1-Y_2|\leqslant c\right)$ 的取值而言, $\mathrm{Var}\left(\omega^{\mathrm{T}}(X_1-X_2)\big||Y_1-Y_2|\leqslant c\right)$ 取值会比较大. 这个线性模型的例子也间接说明了条件方差 $\mathrm{Var}\left(\delta^{\mathrm{T}}(X_1-X_2)\big||Y_1-Y_2|\leqslant c\right)$, $\delta\in\mathbb{R}^p$, $\|\delta\|=1$ 和降维空间 $\mathcal{S}(B)$ 之间的一些内在联系. 进而, Li 等 (2005) 提出如下的等高线回归降维矩阵, 定义为

$$\Lambda_{\mathrm{CR}}(c)=\Sigma_X^{-1/2}E\left[(X_1-X_2)(X_1-X_2)^{\mathrm{T}}\Big||Y_1-Y_2|\leqslant c\right]\Sigma_X^{-1/2},$$

其中 $\Sigma_X^{-1/2}$ 是对称矩阵且有 $\Sigma_X^{-1/2}\Sigma_X^{-1/2}=\Sigma_X^{-1}$. 进一步, 对降维矩阵 $\Lambda_{\mathrm{CR}}(c)$ 进行特征值分解, 有

$$\Lambda_{\mathrm{CR}}(c)\eta_s=\lambda_s\eta_s,\quad\lambda_1\geqslant\lambda_2\geqslant\cdots\geqslant\lambda_{p-1}\geqslant\lambda_p,\quad(8.2.12)$$

其中 η_s 是第 s 个特征值 λ_s 所对应的特征向量. Li 等 (2005) 证明了若协变量 X 服从均值为 μ, 协方差矩阵为 Σ_X 的椭球等高分布, 且降维空间 $\mathcal{S}(B)$ 的结构维数

为 k 时, 由前 k 个最小特征值所对应的特征向量 $\eta_{p-k+1}, \eta_{p-k+2}, \cdots, \eta_p$ 可用于求解降维空间 $\mathcal{S}(B)$, 即有

$$\mathcal{S}\left(\Sigma_X^{1/2}\eta_{p-k+1}, \Sigma_X^{1/2}\eta_{p-k+2}, \cdots, \Sigma_X^{1/2}\eta_p\right) \subseteq \mathcal{S}(B).$$

下面给出等高线回归降维方法的具体算法. 定义降维矩阵 $\Lambda_{\mathrm{CR}}^*(c)$ 如下

$$\Lambda_{\mathrm{CR}}^*(c) = \Sigma_X^{-1/2} E\left[(X_1 - X_2)(X_1 - X_2)^{\mathrm{T}} I\{|Y_1 - Y_2| \leqslant c\}\right] \Sigma_X^{-1/2}.$$

不难看出 $\Lambda_{\mathrm{CR}}^*(c) = \Lambda_{\mathrm{CR}}(c) P(|Y_1 - Y_2| \leqslant c)$. 因此, 对降维矩阵 $\Lambda_{\mathrm{CR}}^*(c)$ 进行特征值分解时, 降维矩阵 $\Lambda_{\mathrm{CR}}^*(c)$ 的特征向量和降维矩阵 $\Lambda_{\mathrm{CR}}(c)$ 的特征向量是相同的.

假设 $\{(X_i, Y_i), i = 1, \cdots, n\}$ 是来自于模型 (8.1.1) 的一组 i.i.d.的随机样本. Li 等 (2005) 给出了降维矩阵 $\Lambda_{\mathrm{CR}}^*(c)$ 的估计量, 具体算法如下.

步骤 1 给出降维矩阵 $\Lambda_{\mathrm{CR}}^*(c)$ 的 U- 统计量估计, 定义如下:

$$\hat{\Lambda}_{\mathrm{CR}}^*(c) = \hat{\Sigma}_X^{-1/2} \hat{H}_n(c) \hat{\Sigma}_X^{-1/2},$$

$$\hat{H}_n(c) = \frac{2}{n(n-1)} \sum_{1 \leqslant i < j \leqslant n} (X_i - X_j)(X_i - X_j)^{\mathrm{T}} I\{|Y_i - Y_j| \leqslant c\},$$

$$\hat{\Sigma}_X = \frac{1}{n}\sum_{i=1}^n \left(X_i - \bar{X}\right)\left(X_i - \bar{X}\right)^{\mathrm{T}}, \quad \bar{X} = \frac{1}{n}\sum_{i=1}^n X_i,$$

其中 $\hat{\Sigma}_X^{-1/2}$ 是对称矩阵且满足 $\hat{\Sigma}_X^{-1/2}\hat{\Sigma}_X^{-1/2} = \hat{\Sigma}_X^{-1}$.

步骤 2 对矩阵 $\hat{\Lambda}_{\mathrm{CR}}^*(c)$ 进行特征值分解, 有

$$\hat{\Lambda}_{\mathrm{CR}}^*(c)\hat{\eta}_s = \hat{\lambda}_s \hat{\eta}_s, \quad \hat{\lambda}_1 \geqslant \hat{\lambda}_2 \geqslant \cdots \geqslant \hat{\lambda}_{p-1} \geqslant \hat{\lambda}_p.$$

步骤 3 将步骤 2 中的前 k 个最小特征值$\hat{\lambda}_{p-k+1}, \hat{\lambda}_{p-k+2}, \cdots, \hat{\lambda}_p$ 所对应的特征向量 $\hat{\eta}_{p-k+1}, \hat{\eta}_{p-k+2}, \cdots, \hat{\eta}_p$ 进行线性变换: $\hat{\Sigma}_X^{1/2}\hat{\eta}_{p-k+1}, \hat{\Sigma}_X^{1/2}\hat{\eta}_{p-k+2}, \cdots, \hat{\Sigma}_X^{1/2}\hat{\eta}_p$, 其中 $\hat{\Sigma}_X^{1/2}$ 是对称矩阵且满足 $\hat{\Sigma}_X^{1/2}\hat{\Sigma}_X^{1/2} = \hat{\Sigma}_X$. 此时, 将变换后的特征向量所张成的空间 $\mathcal{S}\left(\hat{\Sigma}_X^{1/2}\hat{\eta}_{p-k+1}, \hat{\Sigma}_X^{1/2}\hat{\eta}_{p-k+2}, \cdots, \hat{\Sigma}_X^{1/2}\hat{\eta}_p\right)$ 作为降维空间 $\mathcal{S}(B)$ 的估计.

Li 等 (2005) 证明了在一定的正则条件下, $\mathcal{S}\left(\hat{\Sigma}_X^{1/2}\hat{\eta}_{p-k+1}, \hat{\Sigma}_X^{1/2}\hat{\eta}_{p-k+2}, \cdots, \hat{\Sigma}_X^{1/2}\hat{\eta}_p\right)$ 是降维空间 $\mathcal{S}(B)$ 的一个 $\sqrt{n}$ 相合估计.

8.2.4 方向回归方法

在 8.2.3 节中, Li 等 (2005) 注意到在等高线回归降维方法中使用了经验方向向量 $X_i - X_j$ 和条件 $|Y_i - Y_j| \leqslant c$, $1 \leqslant i < j \leqslant n$, 并证明了等高线回归降维矩阵

$\Lambda_{\mathrm{CR}}(c)$ 可以用来求解降维空间 $\mathcal{S}(B)$. 但是降维矩阵 $\Lambda_{\mathrm{CR}}(c)$ 中的常数 c 的选取问题, 目前还没有切实有效的方法. 为克服这一缺点, Li 和 Wang (2007) 在等高线回归降维方法的基础上提出了方向回归 (directional regression, DR) 方法.

定义矩阵

$$V(Y_1,Y_2)=E\left[(X_1-X_2)(X_1-X_2)^{\mathrm{T}}\Big|Y_1,Y_2\right],$$

其中 (X_1,Y_1) 和 (X_2,Y_2) 相互独立, 并且与 (X,Y) 有相同的分布函数. 在矩阵 $V(Y_1,Y_2)$ 的定义中, 不含有任何赘余参数, 只与方向向量 X_1-X_2 和响应变量 Y_1,Y_2 有关系. 注意到矩阵 $V(Y_1,Y_2)$ 可改写为

$$\begin{aligned}V(Y_1,Y_2)=&\operatorname{Cov}(X_1|Y_1)+\operatorname{Cov}(X_2|Y_2)\\&+E\left(X_1|Y_1\right)\left\{E\left(X_1^{\mathrm{T}}|Y_1\right)-E\left(X_2^{\mathrm{T}}|Y_2\right)\right\}\\&+E\left(X_2|Y_2\right)\left\{E\left(X_2^{\mathrm{T}}|Y_2\right)-E\left(X_1^{\mathrm{T}}|Y_1\right)\right\}.\end{aligned}\tag{8.2.13}$$

不难看出, 矩阵 $V(Y_1,Y_2)$ 中含有 $\operatorname{Cov}(X_1|Y_1)$, $\operatorname{Cov}(X_2|Y_2)$, $E\left(X_1|Y_1\right)$ 和 $E(X_2|Y_2)$. 因此基于矩阵 $V(Y_1,Y_2)$ 构造的方向回归降维方法实际上是逆回归方法和逆方差方法的一种组合 (若用切片的估计方法, 可用 SIR 和 SAVE 的估计方法给出). 下面给出方向回归降维方法的降维矩阵 Λ_{DR}, 定义如下:

$$\Lambda_{\mathrm{DR}}=\Sigma_X^{-1}E_{Y_1,Y_2}\left[(2\Sigma_X-V(Y_1,Y_2))(2\Sigma_X-V(Y_1,Y_2))^{\mathrm{T}}\right]\Sigma_X^{-1}.\tag{8.2.14}$$

结合 8.2.1 节和 8.2.2 节中的一些相关推导, 当结构维数 $k=1$ 时, 降维矩阵 Λ_{DR} 可写为

$$\Lambda_{\mathrm{DR}}=c_*\beta_1\beta_1^{\mathrm{T}},$$

其中 $c_*=\left[d(Y_1)+d(Y_2)-(c(Y_1)-c(Y_2))^2\right]^2\beta_1^{\mathrm{T}}\Sigma_X\Sigma_X\beta_1$. 与 8.2.1 节和 8.2.2 节的分析类似, 若记 η_{DR} 是 Λ_{DR} 最大值特征值所对应的特征向量, 那么有 $\eta_{\mathrm{DR}}=\pm\beta_1$, 即 $\mathcal{S}(\eta_{\mathrm{DR}})=\mathcal{S}(\beta_1)$. 若结构维数 $k>1$, Li 和 Wang (2007) 证明了降维矩阵 Λ_{DR} 的前 k 个非零最大特征值 $\lambda_1\geqslant\cdots\geqslant\lambda_k>0$ 所对应的特征向量所张成的空间 $\mathcal{S}(\eta_1,\eta_2,\cdots,\eta_k)$ 满足: $\mathcal{S}(\eta_1,\eta_2,\cdots,\eta_k)\subseteq\mathcal{S}(B)$, 其中 η_s 是特征向量并有 $\Lambda_{\mathrm{DR}}\eta_s=\lambda_s\eta_s$, $s=1,\cdots,k$.

假设 $\{(X_i,Y_i),i=1,\cdots,n\}$ 是来自于模型 (8.1.1) 的一组 i.i.d.的随机样本. 下面给出方向回归降维方法的估计程序, 具体步骤如下.

步骤 1　将 Y 的样本 $\{Y_1,\cdots,Y_n\}$ 按照其取值分为 m 个区间, 分别记为 $J_1,J_2,\cdots,J_m$.

步骤 2 给出降维矩阵 Λ_{DR} 的估计, 定义如下:

$$\hat{\Lambda}_{\mathrm{DR}} = \hat{\Sigma}_X^{-1}\left[2\hat{\Sigma}_X - \hat{V}(J_k, J_l)\right]\left[2\hat{\Sigma}_X - \hat{V}(J_k, J_l)\right]^{\mathrm{T}}\hat{\Sigma}_X^{-1},$$

$$\hat{V}(J_k, J_l) = \frac{\sum\limits_{1\leqslant i<j\leqslant m}(X_i - X_i)(X_i - X_j)^{\mathrm{T}} I\{Y_i \in J_k, Y_j \in J_l\}}{\sum\limits_{1\leqslant i<j\leqslant m} I\{Y_i \in J_k, Y_j \in J_l\}},$$

$$\hat{\Sigma}_X = \frac{1}{n}\sum_{i=1}^{n}\left(X_i - \bar{X}\right)\left(X_i - \bar{X}\right)^{\mathrm{T}}, \qquad \bar{X} = \frac{1}{n}\sum_{i=1}^{n} X_i.$$

步骤 3 对降维矩阵 $\hat{\Lambda}_{\mathrm{DR}}$ 进行特征值分解, 即

$$\hat{\Lambda}_{\mathrm{DR}}\hat{\eta}_s = \hat{\lambda}_s\hat{\eta}_s, \qquad \hat{\lambda}_1 \geqslant \hat{\lambda}_2 \geqslant \cdots \geqslant \hat{\lambda}_k > 0.$$

步骤 4 将步骤 3 中所得到的前 k 个最大特征值 $\hat{\lambda}_1, \hat{\lambda}_2, \cdots, \hat{\lambda}_k$ 所对应的特征向量 $\hat{\eta}_1, \hat{\eta}_2, \cdots, \hat{\eta}_k$ 所张成的空间 $\mathcal{S}(\hat{\eta}_1, \hat{\eta}_2, \cdots, \hat{\eta}_k)$ 作为降维空间 $\mathcal{S}(B)$ 的估计.

8.2.5 Hessian 主方向方法

Hessian 主方向 (PHD) 方法由 Li (1992) 提出, 下面详细介绍 PHD 方法. 首先假设响应变量 Y 和协变量 X 的条件数学期望具有如下的降维结构

$$E(Y|X = x) = g_B(B^{\mathrm{T}}x) =: g_B(v_1, v_2, \cdots, v_k), \tag{8.2.15}$$

其中 $B = (\beta_1, \beta_2, \cdots, \beta_k)$ 是 $p \times k$ 的降维矩阵, $B^{\mathrm{T}}B = I_k$, $v_s = \beta_s^{\mathrm{T}}x$, $s = 1, 2, \cdots, k$. 可以看出, 函数 $g_B(B^{\mathrm{T}}x)$ 的 Hessian 矩阵和降维空间有如下的关系

$$\frac{\partial^2 g_B(B^{\mathrm{T}}x)}{\partial x \partial x^{\mathrm{T}}} = Bg''(x)B^{\mathrm{T}}, \tag{8.2.16}$$

其中 $g''(x)$ 是一个 $k \times k$ 的矩阵, 并且该矩阵的第 (s, l) 个元素为

$$\left.\frac{\partial^2 g_B(v_1, v_2, \cdots, v_k)}{\partial v_s \partial v_l}\right|_{v = B^{\mathrm{T}}x},$$

这里 $v = (v_1, v_2, \cdots, v_k)^{\mathrm{T}}$.

定义 $H_X = BE\left[g''(X)\right]B^{\mathrm{T}}$. 不难看出, 当 $E[g''(X)]$ 是一个非零矩阵时, 若单位向量 e 与降维空间 $\mathcal{S}(B)$ 正交 (即 $e^{\mathrm{T}}B = 0$), 可有 $H_Xe = 0$. 这说明了任何与降维空间 $\mathcal{S}(B)$ 正交的单位向量 e, 都是矩阵 H_X 的零特征值所对应的特征向量. 这也说明了矩阵 H_X 的非零特征值所对应的特征向量可用于求解降维空间 $\mathcal{S}(B)$. PHD 方法的缺陷在于 Hessian 矩阵涉及二阶导数, 使得 PHD 对 $G(u, v) = au + v$

这一特殊的线性结构不甚敏感, 因为此时 Hessian 矩阵是零矩阵, 不存在非零的特征值.

特殊地, 当结构维数 $k=1$ 时, 模型 (8.2.16) 变成了第 7 章所讨论的单指标模型 $Y=g(\beta_1^{\mathrm{T}}X)+\varepsilon$. 此时 $H_X=E(g''(\beta_1^{\mathrm{T}}X))\beta_1\beta_1^{\mathrm{T}}$. 不难看出, 在结构维数等于 1 的情况下, 若记 η_{PHD} 是矩阵 H_X 最大非零特征值所对应的特征向量, 那么有 $\eta_{\mathrm{PHD}}=\pm\beta_1$.

当 X 服从多元正态分布时, Li (1992) 利用 Stein 引理 (Stein, 1981) 给出总体意义下矩阵 H_X 的估计方法, 即

$$H_X=\Sigma_X^{-1}E\left[(Y-E(Y))(X-E(X))(X-E(X))^{\mathrm{T}}\right]\Sigma_X^{-1}. \tag{8.2.17}$$

假设 $\{(X_i,Y_i),i=1,\cdots,n\}$ 是一组来自于模型 (8.1.1) 的一组 i.i.d. 的随机样本. Li (1992) 直接利用矩估计方法给出 H_X 的估计如下:

$$\hat{H}_X=\hat{\Sigma}_X^{-1}\frac{1}{n}\sum_{i=1}^{n}(Y_i-\bar{Y})(X_i-\bar{X})(X_i-\bar{X})^{\mathrm{T}}\hat{\Sigma}_X^{-1}, \tag{8.2.18}$$

其中 $\bar{Y}=\dfrac{1}{n}\sum\limits_{i=1}^{n}Y_i$, $\bar{X}=\dfrac{1}{n}\sum\limits_{i=1}^{n}X_i$, $\hat{\Sigma}_X=\dfrac{1}{n}\sum\limits_{i=1}^{n}\left(X_i-\bar{X}\right)\left(X_i-\bar{X}\right)^{\mathrm{T}}$. 在得到 H_X 的估计 $\hat{H}_X$ 之后, 对 $\hat{H}_X$ 进行如下的特征值分解:

$$\hat{H}_X\hat{\eta}_s=\hat{\lambda}_s\hat{\eta}_s, \qquad |\hat{\lambda}_1|\geqslant|\hat{\lambda}_2|\geqslant\cdots\geqslant|\hat{\lambda}_k|>0. \tag{8.2.19}$$

最终将特征向量 $\hat{\eta}_1,\cdots,\hat{\eta}_k$ 所张成的空间 $\mathcal{S}(\hat{\eta}_1,\hat{\eta}_2,\cdots,\hat{\eta}_k)$ 作为降维空间 $\mathcal{S}(B)$ 的估计. Li (1992) 与 Cook (1998b) 详细讨论了 PHD 降维矩阵 $\hat{H}_X$ 的相关大样本理论性质. 进一步, Lue (2004) 提出了基于 PHD 的测量误差降维方法, Lue (2010) 进一步将 PHD 方法推广到响应变量 Y 是多维的情形.

8.2.6 最小平均方差估计方法

Xia 等 (2002) 考虑了如下具有降维结构的回归模型:

$$Y=g_B(B^{\mathrm{T}}X)+\varepsilon, \tag{8.2.20}$$

其中 B 是维数为 $p\times k$ 的正交矩阵, $B^{\mathrm{T}}B=I_k$, $g_B(\cdot)$ 是未知的 k 元光滑联系函数, 模型误差 ε 满足 $E(\varepsilon|X)=0$. 在第 7 章中, 已经介绍了最小平均方差估计 (MAVE) 方法在部分线性单指标模型中的应用. 本节将介绍降维模型 (8.2.20) 中降维矩阵 B 估计的 MAVE 方法.

模型 (8.2.20) 中的降维矩阵 B, 从总体意义上来看, 可通过极小化如下约束的目标函数得到, 即

$$B=\arg\min_{B_*:B_*^{\mathrm{T}}B_*=I_k}E\left[Y-g_{B_*}(B_*^{\mathrm{T}}X)\right]^2=\arg\min_{B_*:B_*^{\mathrm{T}}B_*=I_k}E\left[\sigma_{B_*}^2(B_*^{\mathrm{T}}X)\right], \tag{8.2.21}$$

其中

$$\sigma_{B_*}^2(t) = E\left[\left(Y - E(Y|B_*^{\mathrm{T}}X)\right)^2 \middle| B_*^{\mathrm{T}}X = t\right]. \tag{8.2.22}$$

因此, Xia 等 (2002) 把这种极小化平均条件方差所得到的估计称为最小平均方差估计, 该方法也称为最小平均方差估计方法, 简写为 MAVE 方法.

记 $B_* = (\beta_{1*}, \beta_{2*}, \cdots, \beta_{k*})$ 为 $p \times k$ 的矩阵, 其中 β_{s*} 为 $p \times 1$ 的列向量, $s = 1, \cdots, k$. 定义函数 $g_{B_*}(v_1, v_2, \cdots, v_k)$ 如下:

$$g_{B_*}(v_1, v_2, \cdots, v_k) =: E\left[Y \middle| \beta_{1*}^{\mathrm{T}}X = v_1, \beta_{2*}^{\mathrm{T}}X = v_2, \cdots, \beta_{k*}^{\mathrm{T}}X = v_k\right].$$

假设 $\{(Y_i, X_i), i = 1, \cdots, n\}$ 是来自于模型 (8.2.20) 的一组 i.i.d. 随机样本. 条件期望 $g_{B_*}(v_1, v_2, \cdots, v_k)$ 可用局部线性展开进行逼近. 对于 X_i 在 x 的任意一个邻域内, 则有

$$\begin{aligned} Y_i - g_{B_*}(B_*^{\mathrm{T}}X_i) &\approx Y_i - g_{B_*}(B_*^{\mathrm{T}}x) - \left[g'_{B_*}(B_*^{\mathrm{T}}x)\right]^{\mathrm{T}} B_*^{\mathrm{T}}(X_i - x) \\ &=: Y_i - a - b^{\mathrm{T}}B_*^{\mathrm{T}}(X_i - x), \end{aligned} \tag{8.2.23}$$

其中 $b = \left(b_{(1)}, b_{(2)}, \cdots, b_{(k)}\right)^{\mathrm{T}}, b_{(s)} = \left.\dfrac{\partial g_{B_*}(v_1, v_2, \cdots, v_k)}{\partial v_s}\right|_{v = B_*^{\mathrm{T}}x}, v = (v_1, v_2, \cdots, v_k)^{\mathrm{T}}$.

进一步, 由 Fan 和 Gijbels (1996) 提出的局部线性光滑的思想可给出 $\sigma_{B_*}^2(B_*^{\mathrm{T}}x)$ 的估计, 定义如下:

$$\hat{\sigma}_{B_*}^2(B_*^{\mathrm{T}}x) = \min_{a,b} \sum_{i=1}^n \left\{Y_i - a - b^{\mathrm{T}}B_*^{\mathrm{T}}(X_i - x)\right\}^2 \omega_{ix}, \tag{8.2.24}$$

其中 $\omega_{ix} \geqslant 0, i = 1, \cdots, n$, 是一些权函数, 满足 $\sum\limits_{i=1}^n \omega_{ix} = 1$.

结合式 (8.2.21) 和式 (8.2.24), 可给出模型 (8.2.20) 中的降维矩阵 B 的估计, 定义如下:

$$\begin{aligned} \hat{B} &= \arg \min_{B_*: B_*^{\mathrm{T}}B_* = I_k} \sum_{j=1}^n \hat{\sigma}_{B_*}^2(B_*^{\mathrm{T}}X_j) \\ &= \arg \min_{\substack{B_*:\, B_*^{\mathrm{T}}B_* = I_k, \\ a_j, b_j, j = 1, \cdots, n}} \sum_{j=1}^n \sum_{i=1}^n \left\{Y_i - a_j - b_j^{\mathrm{T}}B_*^{\mathrm{T}}(X_i - X_j)\right\}^2 \omega_{ij}, \end{aligned} \tag{8.2.25}$$

其中 $\omega_{ij} \geqslant 0, i, j = 1, \cdots, n$, 是一些权函数, 满足 $\sum\limits_{i=1}^n \omega_{ij} = 1$.

MAVE 方法虽然强调了响应变量 Y 和协变量 X 需要具有降维结构, 即满足回归模型 (8.2.20), 但是该方法不依赖于协变量 X 的分布函数, 如不需要满足线性条件和常数方差条件, 同时建立了半参数回归模型和降维模型之间的联系.

8.2.7 条件密度函数的 dMAVE 方法

8.2.6 节介绍的 MAVE 方法是基于响应变量 Y 和协变量 X 之间存在具有降维结构的回归模型 (8.2.20). 若进一步假设给定协变量 X 的条件下, 响应变量 Y 的条件密度函数 $f_{Y|X}(y|x)$ 存在, 那么在降维结构 (8.1.1) 的基础上, 可以得到

$$f_{Y|X}(y|x) = f_{Y|B^{\mathrm{T}}X}(y|B^{\mathrm{T}}x), \tag{8.2.26}$$

其中 $f_{Y|B^{\mathrm{T}}X}(y|B^{\mathrm{T}}x)$ 是给定变量 $B^{\mathrm{T}}X$ 的条件下, 响应变量 Y 的条件密度函数. 从式 (8.2.26), 可直接得到

$$\frac{\partial f_{Y|X}(y|x)}{\partial x} = \frac{\partial f_{Y|B^{\mathrm{T}}X}(y|B^{\mathrm{T}}x)}{\partial B^{\mathrm{T}}x}\frac{\partial B^{\mathrm{T}}x}{\partial x} = B\left(\frac{\partial f_{Y|B^{\mathrm{T}}X}(y|B^{\mathrm{T}}x)}{\partial B^{\mathrm{T}}x}\right). \tag{8.2.27}$$

式 (8.2.27) 明确了条件密度函数 $f_{Y|X}(y|x)$ 关于 x 的偏导数和矩阵 B 之间的关系, 但是条件密度函数 $f_{Y|X}(y|x)$ 通常情况下未知. 为此, Xia (2007) 使用 Fan 等 (1996) 提出的双核局部线性光滑方法来给出条件密度函数的估计. 注意在模型 (8.2.26) 成立的基础上, 有

$$\begin{aligned} f_{Y|X}(y|x) &= \lim_{b\to 0} E\left[H_b(Y-y)\middle|X=x\right] = \lim_{b\to 0} E\left[H_b(Y-y)\middle|B^{\mathrm{T}}X=B^{\mathrm{T}}x\right] \\ &= f_{Y|B^{\mathrm{T}}X}(y|B^{\mathrm{T}}x), \end{aligned} \tag{8.2.28}$$

其中 $H_b(\cdot) = b^{-1}H(\cdot/b)$, $H(\cdot)$ 是核密度函数, b 是窗宽. 在模型 (8.2.26) 的基础上, 将 $H_b(Y-y)$ 看成是新的响应变量, 建立了如下具有降维结构的回归模型

$$H_b(Y-y) = m_b(B^{\mathrm{T}}X, y) + \varepsilon_b(y|X), \tag{8.2.29}$$

其中 $m_b(B^{\mathrm{T}}x, y) = E\left[H_b(Y-y)\middle|B^{\mathrm{T}}X=x\right]$, $\varepsilon_b(y|X) = H_b(Y-y) - E\left[H_b(Y-y)\middle|X\right]$. 这时可将模型 (8.2.29) 看成是 8.2.6 节中降维模型 (8.2.20) 的一种特殊情况. 因此, Xia (2006) 结合降维回归模型 (8.2.29) 给出降维矩阵 B 的两种估计方法: 第 1 种是利用局部线性思想给出偏导数 $\dfrac{m_b(B^{\mathrm{T}}x, y)}{\partial x}$ 的核估计, 再利用该核估计构造统计量用于估计降维空间 $\mathcal{S}(B)$ 的降维矩阵; 第 2 种方法是基于 8.2.6 节介绍的 MAVE 方法给出降维矩阵 B 的估计. 下面首先介绍 Xia (2006) 提出的第 1 种方法.

1. *方法 1*

首先, 对函数 $m_b(B^{\mathrm{T}}x, y)$ 关于 x 求偏导数, 可得到

$$\frac{\partial m_b(B^{\mathrm{T}}x, y)}{\partial x} = B\left(\frac{\partial m_b(v, y)}{\partial v}\bigg|_{v=B^{\mathrm{T}}x}\right) =: Bm_b'(B^{\mathrm{T}}x, y), \tag{8.2.30}$$

其中 $v=(v_1,\cdots,v_k)^{\mathrm{T}}$, $\dfrac{\partial m_b(v,y)}{\partial v}=\left(\dfrac{\partial m_b(v,y)}{\partial v_1},\cdots,\dfrac{\partial m_b(v,y)}{\partial v_k}\right)$. 进一步, 定义矩阵:

$$\Lambda_{\mathrm{D}}=E\left[\left(\left.\frac{\partial m_b(B^{\mathrm{T}}x,y)}{\partial x}\right|_{(x,y)=(X,Y)}\right)\left(\left.\frac{\partial m_b(B^{\mathrm{T}}x,y)}{\partial x}\right|_{(x,y)=(X,Y)}\right)^{\mathrm{T}}\right].$$

由式 (8.2.30), 可得到 $\Lambda_{\mathrm{D}}=BE\left[m_b'(B^{\mathrm{T}}X,Y)m_b'^{\mathrm{T}}(B^{\mathrm{T}}X,Y)\right]B^{\mathrm{T}}$, 因此不难看出矩阵 Λ_{D} 的前 k 个非零特征值所对应的特征向量可用于求解降维空间 $\mathcal{S}(B)$.

假设 $\{(X_i,Y_i),i=1,\cdots,n\}$ 是来自于模型 (8.2.26) 的一组 i.i.d.的随机样本. Xia (2007) 利用回归模型 (8.2.29) 和局部线性光滑的思想, 给出偏导数 $\dfrac{\partial m_b(B^{\mathrm{T}}x,y)}{\partial x}$ 的估计, 并用矩估计方法直接给出矩阵 Λ_{D} 的估计, 定义如下:

$$\begin{aligned}\hat{\Lambda}_{\mathrm{D}}=\frac{1}{n^2}\sum_{s=1}^{n}\sum_{k=1}^{n}&\left(\left.\frac{\widehat{\partial m_b}(B^{\mathrm{T}}x,y)}{\partial x}\right|_{(x,y)=(X_s,Y_k)}\right)\\&\times\left(\left.\frac{\widehat{\partial m_b}(B^{\mathrm{T}}x,y)}{\partial x}\right|_{(x,y)=(X_s,Y_k)}\right)^{\mathrm{T}},\end{aligned}\tag{8.2.31}$$

其中 $\left.\dfrac{\widehat{\partial m_b}(B^{\mathrm{T}}x,y)}{\partial x}\right|_{(x,y)=(X_s,Y_k)}=\hat{b}_{sk}$, 且 $\hat{b}_{sk}$ 可以通过极小化下面的局部加权目标函数而得到, 即

$$\begin{aligned}&(\hat{a}_{sk},\hat{b}_{sk})\\&=\arg\min_{a_{sk},b_{sk}}\sum_{i=1}^{n}\left\{H_b(Y_i-Y_k)-a_{sk}-b_{sk}^{\mathrm{T}}(X_i-X_s)\right\}^2K_h(X_i-X_s),\end{aligned}\tag{8.2.32}$$

其中 $K_h(X_i-X_s)=\prod_{j=1}^{p}h_j^{-1}K_j\left(\dfrac{X_{ij}-X_{sj}}{h_j}\right)$, $K_j(\cdot)$ 是核密度函数, h_j 是窗宽. 在式 (8.2.32) 中, 使用了下面的局部线性逼近:

$$m_b(B^{\mathrm{T}}X_i,Y_k)\approx m_b(B^{\mathrm{T}}X_s,Y_k)+(Bm_b'(B^{\mathrm{T}}X_s,Y_k))^{\mathrm{T}}(X_i-X_s)=:a_{sk}+b_{sk}^{\mathrm{T}}(X_i-X_j).$$

在得到矩阵 Λ_{D} 的估计 $\hat{\Lambda}_{\mathrm{D}}$ 后, 可对矩阵 $\hat{\Lambda}_{\mathrm{D}}$ 进行特征值分解, 若记 $\hat{\eta}_1,\hat{\eta}_2,\cdots,\hat{\eta}_k$ 分别是 $\hat{\Lambda}_{\mathrm{D}}$ 的前 k 个非零最大特征值所对应的特征向量, 则特征向量 $\hat{\eta}_1,\hat{\eta}_2,\cdots,\hat{\eta}_k$ 所张成的空间 $\mathcal{S}(\hat{\eta}_1,\hat{\eta}_2,\cdots,\hat{\eta}_k)$ 是降维空间 $\mathcal{S}(B)$ 的相合估计. Xia (2007) 详细讨论了估计 $\mathcal{S}(\hat{\eta}_1,\hat{\eta}_2,\cdots,\hat{\eta}_k)$ 的相关理论性质.

2. *方法* 2

对于第 1 种方法, 主要是把 $B\dfrac{\partial m_b(v,y)}{\partial v}\Big|_{v=B^{\mathrm{T}}x}$(或 $Bm_b'(B^{\mathrm{T}}x,y)$) 作为一个整体进行估计. 同时, 在式 (8.2.32) 中, 需要使用 p 维核函数:

$$K_h(X_i-X_s)=\prod_{j=1}^{p}h_j^{-1}K_j\left(\frac{X_{ij}-x_{sj}}{h_j}\right)$$

进行局部光滑. 当维数 p 过大时, "维数祸根" 会使得第 1 种方法的估计效率大大降低. 为了克服这个问题, 基于 MAVE 方法的思想, 提出极小化下面的目标函数:

$$\sum_{k=1}^{n}\sum_{s=1}^{n}\sum_{i=1}^{n}\left\{H_b(Y_i-Y_k)-a_{sk}-c_{sk}^{\mathrm{T}}B^{\mathrm{T}}(X_i-X_s)\right\}^2K_h(B^{\mathrm{T}}X_i-B^{\mathrm{T}}X_s). \quad (8.2.33)$$

在式 (8.2.33) 中, 使用了下面的局部线性逼近:

$$m_b(B^{\mathrm{T}}X_i,Y_k)\approx m_b(B^{\mathrm{T}}X_s,Y_k)+{m'}_b^{\mathrm{T}}(B^{\mathrm{T}}X_s,Y_k)B^{\mathrm{T}}(X_i-X_s)=:a_{sk}+c_{sk}^{\mathrm{T}}B^{\mathrm{T}}(X_i-X_j),$$

其中 $K_h(B^{\mathrm{T}}X_i-B^{\mathrm{T}}X_j)=\prod\limits_{j=1}^{k}h_j^{-1}K_j\left(\dfrac{\beta_j^{\mathrm{T}}X_{ij}-\beta_j^{\mathrm{T}}X_{sj}}{h_j}\right)$, $K_j(\cdot)$ 是核密度函数, h_j 是窗宽. 不难看出, 若结构维数 $k\leqslant 3$ 时, 关于目标函数 (8.2.26) 的极小化过程就在一定程度上避免了 "维数祸根" 问题, 也提高了估计效率. 在目标函数 (8.2.33) 中, 对未知量 a_{sk},c_{sk} 和 B 估计, 可采用类似于 7.1 节中的迭代算法.

8.2.8 累积分布降维方法

dMAVE 方法先将条件密度函数 (8.2.26) 用回归方式表示出来, 再结合局部线性光滑方法给出条件密度函数中降维空间 $\mathcal{S}(B)$ 的估计. 若从条件密度函数出发可以给出降维空间的估计, Zhu 等 (2010) 指出降维空间同样可以通过条件分布的方法给出. 若进一步假设给定协变量 X 的条件下, 响应变量 Y 的条件分布 $F_{Y|X}(y|x)$ 存在, 那么在降维模型 (8.1.1) 的基础上, 可以得到

$$F_{Y|X}(y|x)=F_{Y|B^{\mathrm{T}}X}(y|B^{\mathrm{T}}x), \quad (8.2.34)$$

其中 $F_{Y|B^{\mathrm{T}}X}(y|B^{\mathrm{T}}x)$ 是给定变量 $B^{\mathrm{T}}X$ 的条件下, 响应变量 Y 的条件分布函数. 由式 (8.2.34), 可直接得到

$$\frac{\partial F_{Y|X}(y|x)}{\partial x}=\frac{\partial F_{Y|B^{\mathrm{T}}X}(y|B^{\mathrm{T}}x)}{\partial B^{\mathrm{T}}x}\frac{\partial B^{\mathrm{T}}x}{\partial x}=B\left(\frac{\partial F_{Y|B^{\mathrm{T}}X}(y|B^{\mathrm{T}}x)}{\partial B^{\mathrm{T}}x}\right). \quad (8.2.35)$$

式 (8.2.35) 明确了条件分布函数 $F_{Y|X}(y|x)$ 关于 x 的偏导数和降维矩阵 B 之间的关系.

此外, 降维模型 (8.1.1) 保证了下面恒等式成立:

$$\begin{aligned}F_{Y|X}(y|x) &= E[I\{Y \leqslant y\}|X = x] = E\left[I\{Y \leqslant y\}|B^{\mathrm{T}}X = B^{\mathrm{T}}x\right] \\ &= F_{Y|B^{\mathrm{T}}X}(y|B^{\mathrm{T}}x).\end{aligned} \tag{8.2.36}$$

不妨假设 $E(X) = 0$, 结合式 (8.2.36) 和线性条件 (8.2.1), Zhu 等 (2010) 推得

$$\begin{aligned}\Sigma_X^{-1}E[XI\{Y \leqslant y\}] &= \Sigma_X^{-1}E[XE(I\{Y \leqslant y\}|X)] = \Sigma_X^{-1}E[XF_{Y|X}(y|X)] \\ &= \Sigma_X^{-1}E[XF_{Y|B^{\mathrm{T}}X}(y|B^{\mathrm{T}}X)] \\ &= \Sigma_X^{-1}E\left[F_{Y|B^{\mathrm{T}}X}(y|B^{\mathrm{T}}X)E\left(X\middle|B^{\mathrm{T}}X\right)\right] \\ &= B\left(B^{\mathrm{T}}\Sigma_X B\right)^{-1}E\left[B^{\mathrm{T}}XF_{Y|B^{\mathrm{T}}X}(y|B^{\mathrm{T}}X)\right].\end{aligned} \tag{8.2.37}$$

式 (8.2.37) 表明 $\Sigma_X^{-1}E[XI\{Y \leqslant y\}] \subseteq \mathcal{S}(B)$.

若记 $M(y) = E[XI\{Y \leqslant y\}]$, Zhu 等 (2010) 利用式 (8.2.37) 构造了如下的降维矩阵:

$$\Lambda_{\mathrm{CU}} = \Sigma_X^{-1}E\left[M(Y^*)M^{\mathrm{T}}(Y^*)\right]\Sigma_X^{-1}, \tag{8.2.38}$$

其中 Y^* 是与响应变量 Y 相互独立, 且有相同的分布函数.

特殊地, 当结构维数 $k = 1$(即 $B = \beta_1$) 时, 降维矩阵 (8.2.38) 可简化为

$$\begin{aligned}\Lambda_{\mathrm{CU}} &= E\left[m^2(Y^*)\right]\beta_1\beta_1^{\mathrm{T}}, \\ m(Y^*) &= \frac{E_{(X,Y)}\left[\beta_1^{\mathrm{T}}XF_{Y|\beta_1^{\mathrm{T}}X}(Y^*|\beta_1^{\mathrm{T}}X)\right]}{\beta_1^{\mathrm{T}}\Sigma_X\beta_1},\end{aligned}$$

其中 $E_{(X,Y)}(\cdot)$ 表示对随机向量 (X, Y) 求数学期望.

假设 $\{(X_i, Y_i), i = 1, \cdots, n\}$ 是一组来自于模型 (8.2.34) 的一组 i.i.d.的随机样本. Zhu 等 (2010) 给出了式 (8.2.38) 的矩估计, 定义如下:

$$\hat{\Lambda}_{\mathrm{CU}} = \hat{\Sigma}_X^{-1}\frac{1}{n}\sum_{i=1}^{n}\hat{M}(Y_i)\hat{M}^{\mathrm{T}}(Y_i)\hat{\Sigma}_X^{-1}, \quad \hat{M}(Y_i) = \frac{1}{n}\sum_{j=1}^{n}X_jI\{Y_j \leqslant Y_i\},$$

$$\hat{\Sigma}_X = \frac{1}{n}\sum_{i=1}^{n}\left(X_i - \bar{X}\right)\left(X_i - \bar{X}\right)^{\mathrm{T}}, \quad \bar{X} = \frac{1}{n}\sum_{i=1}^{n}X_i.$$

得到降维矩阵的估计 $\hat{\Lambda}_{\mathrm{CU}}$ 以后, 可对降维矩阵 $\hat{\Lambda}_{\mathrm{CU}}$ 进行特征值分解, 即

$$\hat{\Lambda}_{\mathrm{CU}}\hat{\eta}_s = \hat{\lambda}_s\hat{\eta}_s, \qquad \hat{\lambda}_1 \geqslant \hat{\lambda}_2 \geqslant \cdots \geqslant \hat{\lambda}_k > 0.$$

最后, 将前 k 个最大特征值 $\hat{\lambda}_1, \hat{\lambda}_2, \cdots, \hat{\lambda}_k$ 所对应的特征向量 $\hat{\eta}_1, \hat{\eta}_2, \cdots, \hat{\eta}_k$ 所张成的空间 $\mathcal{S}(\hat{\eta}_1, \hat{\eta}_2, \cdots, \hat{\eta}_k)$ 作为降维空间 $\mathcal{S}(B)$ 的估计. Zhu 等 (2010) 进一步讨论了协变量 X 的维数在发散情形下, 估计量 $\hat{\Lambda}_{\mathrm{CU}}$ 的大样本性质.

8.2.9 结构维数的选取

在实际问题中, 一旦结构维数 k 确定, 那么就可以使用前面提到的各种降维方法来给出降维空间 $\mathcal{S}(B)$ 的估计. 但是在实际问题中, 结构维数 k 的大小一般是未知的, 需要发展数据驱动的方法进行选取, 本节讨论一些常用的结构维数 k 的选取方法. 在本节的讨论中, 不妨以逆回归降维方法, 即式 (8.2.5) 所定义的逆回归矩阵 Λ_{IR} 为例进行说明.

1. BIC *准则*

Zhu 等 (2006) 将 Zhao 等 (1986a, 1986b) 用于模型选择的 BIC 准则应用于降维矩阵结构维数的选取问题. 记 $\hat{\Lambda}_{\mathrm{IR}}$ 是降维矩阵 Λ_{IR} 的一个相合估计 (如切片逆回归估计或者核逆回归估计), $\hat{\lambda}_1, \hat{\lambda}_2, \cdots, \hat{\lambda}_p$ 是矩阵 $\hat{\Lambda}_{\mathrm{IR}} + I_p$ 的特征值. 记 $\hat{\tau} = \#\{\hat{\lambda}_1 > 1, \hat{\lambda}_2 > 1, \cdots, \hat{\lambda}_p > 1\}$, 即 $\hat{\tau}$ 是特征值 $\{\hat{\lambda}_s, s = 1, \cdots, p\}$ 中大于 1 的个数. 定义

$$
\begin{aligned}
\log(L_s) &= \frac{n}{2} \sum_{i=1+\min\{\hat{\tau}, s\}}^{p} \left(\log(\hat{\lambda}_i) + 1 - \hat{\lambda}_i\right), \\
G(s) &= \log(L_s) - C_n \frac{s(2p-s+1)}{2},
\end{aligned}
$$

其中 C_n 是一个大于 0 的惩罚参数, $\dfrac{s(2p-s+1)}{2}$ 是自由度. 该自由度的解释如下, 若矩阵 Λ_{IR} 有 s 个非零特征根, 即矩阵 Λ_{IR} 的秩为 s. 不妨假设矩阵 Λ_{IR} 剩余的 $p-s$ 列向量可由前 s 列向量线性表示, 注意到前 s 列列向量总共有 ps 个分量. 再由于矩阵 Λ_{IR} 是对称矩阵, 前 s 列列向量中有 $\dfrac{s(s-1)}{2}$ 个分量相同, 那么前 s 列列向量中可自由变动的分量个数为 $sp - \dfrac{s(s-1)}{2} = \dfrac{s(2p-s+1)}{2}$.

最后可极大化下面的 BIC 准则, 得到结构维数 k 的估计 $\hat{k}$, 定义如下:

$$
\hat{k} = \arg \max_{0 \leqslant s \leqslant p-1} G(s). \tag{8.2.39}
$$

Zhu 等 (2006) 证明了由 BIC 准则选取的结构维数 $\hat{k}$ 依概率 1 收敛到真实的结构维数 k.

2. Bootstrap *方法*

Ye 和 Weiss (2003) 提出基于 bootstrap 重抽样的方法来给出结构维数 k 的选取问题. 假设结构维数的真值为 s, 可根据 $\hat{\Lambda}_{\mathrm{IR}}$ 的前 s 个非零特征值所对应的特征向量给出降维空间 $\mathcal{S}(B)$ 的估计, 不妨记该估计为 $\mathcal{S}(\hat{\eta}_1, \hat{\eta}_2, \cdots, \hat{\eta}_s)$. 然后利

用 Efron (1979) 提出的 bootstrap 重抽样技术重复进行 M 次抽样, 利用这 M 组数据得到 M 个降维矩阵的估计, 记为 $\hat{\Lambda}_{\rm IR}^{(m)}$, 以及相应的 M 个降维空间, 记为 $\mathcal{S}\left(\hat{\eta}_1^{(m)},\hat{\eta}_2^{(m)},\cdots,\hat{\eta}_s^{(m)}\right)$, $m=1,\cdots,M$, 这里的 $\left\{\hat{\eta}_1^{(m)},\hat{\eta}_2^{(m)},\cdots,\hat{\eta}_s^{(m)}\right\}$ 是降维矩阵 $\hat{\Lambda}_{\rm IR}^{(m)}$ 的特征向量.

为了给出结构维数 k 的估计, Ye 和 Weiss (2003) 的 bootstrap 方法的核心思想是: 使用 bootstrap 方法得到的降维空间 $\mathcal{S}\left(\hat{\eta}_1^{(m)},\hat{\eta}_2^{(m)},\cdots,\hat{\eta}_s^{(m)}\right)$ 和原始样本得到的降维空间 $\mathcal{S}(\hat{\eta}_1,\hat{\eta}_2,\cdots,\hat{\eta}_s)$ 之间距离的波动性应该比较小.

记 A_l 是 $p\times t$ 的矩阵, 且满足 $A_l^{\rm T}A_l=I_t$, 其中 I_t 是 $t\times t$ 的单位矩阵, $l=1,2$. 并记由 A_l 的列向量所张成的线性空间为 $\mathcal{S}(A_l)$. 那么空间 $\mathcal{S}(A_1)$ 和空间 $\mathcal{S}(A_2)$ 之间的距离定义如下:

$$d_{\rm CE}(t)=\sqrt{\prod_{i=1}^{t}\lambda_i^2},\qquad d_{\rm TR}(t)=\sqrt{\frac{1}{t}\sum_{i=1}^{t}\lambda_i^2},$$

其中 λ_i^2 是矩阵 $A_2^{\rm T}A_1A_1^{\rm T}A_2$ 的特征值, 且满足 $1\geqslant\lambda_1^2\geqslant\lambda_2^2\geqslant\cdots\geqslant\lambda_t^2\geqslant 0$. 实际上, 距离函数 $d_{\rm CE}(t)$ 是向量之间的相关系数 (Hotelling, 1936), 距离函数 $d_{\rm TR}(t)$ 是向量之间的迹相关系数 (Hooper, 1959). 距离函数 $d_{\rm CE}(t)$ 或者 $d_{\rm TR}(t)$ 的取值在 0 和 1 之间, 取值越接近 1, 表明空间 $\mathcal{S}(A_1)$ 和 $\mathcal{S}(A_2)$ 之间的距离越近.

Ye 和 Weiss (2003)使用距离函数 $d_{\rm CE}(t)$ 和距离函数 $d_{\rm TR}(t)$ 来考察 boostrap 抽样之后的降维空间估计 $\mathcal{S}\left(\hat{\eta}_1^{(m)},\hat{\eta}_2^{(m)},\cdots,\hat{\eta}_s^{(m)}\right)$ 和原始数据得到的降维空间估计 $\mathcal{S}(\hat{\eta}_1,\ \hat{\eta}_2,\cdots,\hat{\eta}_s)$ 之间的距离, 分别可得 M 个距离函数的值: $d_{\rm CE}^{(m)}(t)$ 和 $d_{\rm TR}^{(m)}(t)$, $m=1,\cdots,M$. 记 $\bar{d}_{\rm CE}(t)=\dfrac{1}{M}\sum\limits_{m=1}^{M}d_{\rm TR}^{(m)}(t)$ 和 $\bar{d}_{\rm TR}(t)=\dfrac{1}{M}\sum\limits_{m=1}^{M}d_{\rm TR}^{(m)}(t)$. 最后, 降维空间结构维数的估计定义为

$$\hat{k}=\arg\max_{0\leqslant t\leqslant p-1}\bar{d}_{\rm CE}(t)\quad\text{或者}\quad\hat{k}=\arg\max_{0\leqslant t\leqslant p-1}\bar{d}_{\rm TR}(t).$$

Ye 和 Weiss (2003) 通过数值模拟说明了 boostrap 方法的有效性.

3. 特征值比值方法

Luo 等 (2009) 提出了利用特征值比值的方法来研究结构维数的选取问题. 若降维矩阵 $\Lambda_{\rm IR}$ 有 k 个非零特征值 $\{\lambda_1,\lambda_2,\cdots,\lambda_k\}$ 和 $p-k$ 个零特征值, 不妨记为 $\lambda_1\geqslant\lambda_2\geqslant\cdots\geqslant\lambda_k>0=\lambda_{k+1}=\cdots=\lambda_p$. 记 $\{\hat{\lambda}_1,\hat{\lambda}_2,\cdots,\hat{\lambda}_p\}$ 为降维矩阵估计 $\hat{\Lambda}_{\rm IR}$ 的特征值, 且有 $\hat{\lambda}_1\geqslant\hat{\lambda}_2\geqslant\cdots\geqslant\hat{\lambda}_p\geqslant 0$. 若估计 $\hat{\Lambda}_{\rm IR}$ 满足 $\sqrt{n}\left(\hat{\Lambda}_{\rm IR}-\Lambda_{\rm IR}\right)$ 依分布收敛到正态分布, 那么 Eaton 和 Tyler (1994) 研究表明 $\hat{\lambda}_j$ 的收敛速度相同, $j=1,\cdots,p$. 直观上有: $\dfrac{\hat{\lambda}_j}{\hat{\lambda}_{j+1}}=O_P(1)$, $j=1,\cdots,k-1$. 而 $\dfrac{\hat{\lambda}_k}{\hat{\lambda}_{k+1}}\xrightarrow{P}\infty$. 基于此

想法, Luo 等 (2009) 定义降维空间结构维数的估计为

$$\hat{k} = \arg\min_{1 \leqslant j \leqslant d_{\max}} \left\{ \frac{\hat{\lambda}_j}{\hat{\lambda}_{j+1}} \right\},$$

其中 $d_{\max}$ 是一个预先给定的值, Luo 等 (2009) 建议在实际中可选取 $d_{\max} = 5$.

4. *序贯检验方法*

序贯检验方法来确定结构维数 k 通常是基于降维矩阵估计的特征值的渐近分布来完成的. 不妨以 SIR 的降维矩阵估计 $\hat{\Lambda}_{\mathrm{SIR}}$ 为例, 记 $\hat{\lambda}_s (s = 1, \cdots, p)$ 是 $\hat{\Lambda}_{\mathrm{SIR}}$ 的特征值, 并有 $\hat{\lambda}_1 \geqslant \hat{\lambda}_2 \geqslant \cdots \geqslant \hat{\lambda}_p \geqslant 0$. 若 X 服从正态分布, Li (1991) 证明了 $n \sum_{i=0}^{p-k-1} \hat{\lambda}_{p-i}$ 依分布收敛到自由度为 $(p-k)(H-k-1)$ 的标准 χ^2 分布, 其中 H 为降维矩阵估计 $\hat{\Lambda}_{\mathrm{SIR}}$ 中使用的切片个数. 那么对 SIR 方法而言, 结构维数的确定可通过一系列的假设检验问题进行确定:

$$H_0^{(j)} : k = j \longleftrightarrow H_1^{(j)} : k > j,$$

其中 $j = 0, 1, \cdots, p-1$. 定义统计量在原假设 $H_0^{(j)}$ 成立时的 p 值为

$$p_j = P\left(\chi^2_{(p-j)(H-j-1)} \geqslant n \sum_{i=0}^{p-j-1} \hat{\lambda}_{p-j} \right).$$

若取检验水平为 $\alpha \in (0,1)$, 容易得到, 当 $p_j \leqslant \alpha$ 时, 结构维数 k 至少是 $j+1$. 该假设检验从 $j = 0$ 开始, 若当 $s = 0, 1, \cdots, t$, $p_s \leqslant \alpha$, 且 $p_{t+1} > \alpha$ 时, 则此时结构维数 k 的估计为 t.

序贯检验的方法可适用于不同的降维方法, 但需要确定出降维矩阵特征值估计的渐近分布. 关于该方法的研究, 已经有一些理论结果可供参考, 见 Schott (1994), Bura 和 Cook (2001a, 2001b) 以及 Zeng (2008).

8.3　可加测量误差降维模型

考虑如下的测量误差降维模型:

$$\begin{cases} Y \perp\!\!\!\perp X | B^{\mathrm{T}} X, & \perp\!\!\!\perp \text{ 表示独立}, \\ W = \gamma + \Gamma X + \delta, \end{cases} \tag{8.3.1}$$

其中 Y 是 1 维的响应变量, $X = (X_1, \cdots, X_p)^{\mathrm{T}}$ 是 p 维不可观测的协变量, $B = (\beta_1, \cdots, \beta_k)$ 是一个 $p \times k$ 维的未知列满秩常数降维矩阵. 假定降维矩阵 B 满足

条件 $B^{\mathrm{T}}B=I_k$, 其中 I_k 是 $k\times k$ 维单位矩阵, 即降维矩阵 B 的列向量相互正交, $\|\beta_s\|=1$, $s=1,\cdots,k$, $\beta_s^{\mathrm{T}}\beta_t=0$, $1\leqslant s\neq t\leqslant k$. 模型 (8.3.1) 中的 γ 是 $r\times 1$ 维的常数向量, Γ 是 $r\times p$ 常数矩阵, 并且有 $r\geqslant p$, δ 是一个 r 维的误差向量, 并且和 X 相互独立. W 是可观测的 $r\times 1$ 维的协变量. 本节关心的是如何利用观测到的 (Y,W) 来求解降维模型 (8.3.1) 中由矩阵 B 的列向量所张成的降维空间 $\mathcal{S}(B)$ 的估计问题.

8.3.1 最小二乘方法

本节考虑如何用传统的最小二乘方法关于 (Y,W) 建立回归关系, 来给出降维空间 $\mathcal{S}(B)$ 的估计. Corroll 和 Li (1992) 讨论了结构维数 $k=1$ 情形下的最小二乘估计方法和降维空间之间的联系. 在结构维数 $k=1$ 这一特殊情况下 (即 $B=\beta_1$), 不妨假设 $E(X)=0$, 线性条件 (8.2.1) 可简化为

$$E\left(X|\beta^{\mathrm{T}}X\right)=\frac{\Sigma_X\beta_1\beta_1^{\mathrm{T}}X}{\beta_1^{\mathrm{T}}\Sigma_X\beta_1}. \tag{8.3.2}$$

在线性条件 (8.3.2) 成立的情形下, Carroll 和 Li (1992) 得到了如下的定理.

定理 8.3.1 假设 $E(W)=0, E(X)=0$, 并且当模型假设(8.1.1)和线性条件(8.3.2) 成立时, 则有

$$E(L^{\mathrm{T}}W|Y)=\frac{E(\beta_1^{\mathrm{T}}X|Y)}{\beta_1^{\mathrm{T}}\Sigma_X\beta_1}L^{\mathrm{T}}\Gamma\Sigma_X\beta_1=\frac{E(\beta_1^{\mathrm{T}}X|Y)}{\beta_1^{\mathrm{T}}\Sigma_X\beta_1}L^{\mathrm{T}}\Sigma_W L\beta_1,$$

其中 $L=\Sigma_W^{-1}\mathrm{Cov}(W,X)$, $\Sigma_W=\mathrm{Cov}(W,W)$.

定理 8.3.1 揭示了条件数学期望 $E(L^{\mathrm{T}}W|Y)$ 是 β_1 的线性函数. 换句话说, 如果进一步假设 $L^{\mathrm{T}}\Sigma_W L$ 的逆矩阵存在, 那么 $\left(L^{\mathrm{T}}\Sigma_W L\right)^{-1}E(L^{\mathrm{T}}W|Y)$ 与 β_1 成比例. 与 8.2.1 节介绍的逆回归方法进行比较, 不难看出, 若 L 已知, 可将 $L^{\mathrm{T}}W$ 看成是新的协变量, 可直接将 8.2 节中的降维方法应用于变换后的变量 $(L^{\mathrm{T}}W,Y)$ 来给出降维空间 $\mathcal{S}(B)$ 的估计. 由此想法, Carroll 和 Li (1992) 基于替换变量 $V=L^{\mathrm{T}}W$, 提出了总体意义下的最小二乘估计方法, 即

$$(a,b)=\arg\min_{a_0\in\mathbb{R},b_0\in\mathbb{R}^p}E\left[Y-a_0-b_0^{\mathrm{T}}V\right]^2. \tag{8.3.3}$$

注意到 $\Sigma_V=\mathrm{Cov}(V,V)=L^{\mathrm{T}}\Sigma_W L$. 不难证明: 在定理 8.3.1 的假设条件下, 由式 (8.3.3) 得到的 b 和 β_1 有下面的关系:

$$b=\Sigma_V^{-1}E(YV)=c_{\beta_1}\Sigma_V^{-1}L^{\mathrm{T}}\Sigma_W L\beta_1=c_{\beta_1}\beta_1, \tag{8.3.4}$$

其中 $c_{\beta_1}=\dfrac{E(Y\beta_1^{\mathrm{T}}X)}{\beta_1^{\mathrm{T}}\Sigma_X\beta_1}$.

由 $\|\beta_1\|=1$, 不难看出由最小二乘方法确定的 b 和 β_1 满足: $\beta_1=\pm\frac{b}{\|b\|}$, 即 $\mathcal{S}(b)=\mathcal{S}(\beta_1)$. 因此, Carroll 和 Li (1992) 提出了下面三种情形下的最小二乘估计.

1. L 已知的情形

假设 $\{(W_i,Y_i),i=1,\cdots,n\}$ 是来自于模型 (8.3.1) 的一组 i.i.d. 的随机样本. 对数据 $\{W_1,\cdots,W_n\}$ 作变换: $V_i=L^{\mathrm{T}}W_i$, $i=1,\cdots,n$. 关于数据集 $\{(Y_i,V_i),i=1,\cdots,n\}$ 使用最小二乘方法直接得到 b 的估计 $\hat{b}_{ls}$, 定义如下

$$(\hat{a}_{ls},\hat{b}_{ls})=\arg\min_{a_0\in\mathbb{R},b_0\in\mathbb{R}^p}\sum_{i=1}^{n}E\left[Y_i-a_0-b_0^{\mathrm{T}}V_i\right]^2.$$

进一步计算可得

$$\hat{b}_{ls}=\hat{\Sigma}_V^{-1}\frac{1}{n}\sum_{i=1}^{n}Y_i\left(V_i-\bar{V}\right), \tag{8.3.5}$$

其中 $\hat{\Sigma}_V=\frac{1}{n}\sum_{i=1}^{n}\left(V_i-\bar{V}\right)\left(V_i-\bar{V}\right)^{\mathrm{T}}$, $\bar{V}=\frac{1}{n}\sum_{i=1}^{n}V_i$. Carroll 和 Li (1992) 得到了估计 $\hat{b}_{ls}$ 的渐近表达式为

$$\hat{b}_{ls}=b+\Sigma_V^{-1}\frac{1}{n}\sum_{i=1}^{n}\left(V_i-E(V)\right)e_i+O_P(n^{-1}),$$

其中 $e_i=Y_i-E(Y)-b^{\mathrm{T}}\left(V_i-E(V)\right)$.

2. L 未知, 利用核实数据估计的情形

若有关于变量 (W,X) 的独立于 $\{(Y_i,W_i),i=1,\cdots,n\}$ 的 i.i.d. 核实数据 $\{(\widetilde{W}_j,\widetilde{X}_j),\ j=1,\cdots,m\}$. 不妨假设核实数据的样本数 m 小于 n, Carroll 和 Li (1992) 建立了变量 (W,X) 之间的回归关系 $X=L^{\mathrm{T}}W+\eta, E(\eta)=0$, 并得到 L 的最小二乘估计如下:

$$\hat{L}_{\mathrm{LS}}=\hat{\Sigma}_{\widetilde{W}}^{-1}\frac{1}{n}\sum_{j=1}^{m}\widetilde{W}_j\left(\widetilde{X}_j-\overline{\widetilde{X}}\right)^{\mathrm{T}}, \tag{8.3.6}$$

其中 $\hat{\Sigma}_{\widetilde{W}}=\frac{1}{m}\sum_{j=1}^{m}\left(\widetilde{W}_j-\overline{\widetilde{W}}\right)\left(\widetilde{W}_j-\overline{\widetilde{W}}\right)^{\mathrm{T}}$, $\overline{\widetilde{W}}=\frac{1}{m}\sum_{j=1}^{m}\widetilde{W}_j$. 这时直接使用估计量 $\hat{L}_{\mathrm{LS}}$ 做变换 $\hat{V}_{\mathrm{LS},i}=\hat{L}_{\mathrm{LS}}^{\mathrm{T}}W_i$, $i=1,\cdots,n$. 类似地, 使用最小二乘方法直接得到 b 的估计 $\widetilde{b}_{\mathrm{LS}}$ 如下

$$\widetilde{b}_{\mathrm{LS}}=\hat{\Sigma}_{V*}^{-1}\frac{1}{n}\sum_{i=1}^{n}Y_i\left(\hat{V}_{\mathrm{LS},i}-\overline{\hat{V}}_{\mathrm{LS}}\right), \tag{8.3.7}$$

其中 $\hat{\Sigma}_{V*} = \frac{1}{n}\sum_{i=1}^{n}\left(\hat{V}_{\mathrm{LS},i} - \overline{\hat{V}}_{\mathrm{LS}}\right)\left(\hat{V}_{\mathrm{LS},i} - \overline{\hat{V}}_{\mathrm{LS}}\right)^{\mathrm{T}}, \overline{\hat{V}}_{\mathrm{LS}} = \frac{1}{n}\sum_{i=1}^{n}\hat{V}_{\mathrm{LS},i}$.

3. L 未知, 利用重复测量数据估计的情形

在很多实际问题中, 对某项指标重复测量数次可得到重复测量数据. 跟核实数据相比, 重复测量数据在实际中更容易获得. 在这种情形下, 式 (8.1.2) 中 $\Gamma = I_p$, 则模型 (8.1.2) 即为传统意义上的重复测量误差数据模型. 不妨假设重复观测的次数都相同, 即

$$W_{lj}^* = X_l^* + \delta_{lj}, \quad j = 1, 2, \quad l = 1, \cdots, m, \tag{8.3.8}$$

其中 $\{\delta_{lj}, j = 1, 2, l = 1, \cdots, m\}$ 是同分布且独立于 $\{(W_i, Y_i), i = 1, \cdots, n\}$ 的测量误差. 计算可得

$$\mathrm{Cov}(W_{l1}^* - W_{l2}^*) = 2\mathrm{Cov}(\delta, \delta) = 2\Sigma_\delta,$$
$$\mathrm{Cov}(W_{l1}^* + W_{l2}'^*) = 4\Sigma_X + 2\Sigma_\delta.$$

因此, 可用样本 $\{W_{lj}^*, j = 1, 2, l = 1, \cdots, m\}$ 来估计 Σ_δ 和 Σ_{W^*}, 此时将估计分别记为 $\hat{\Sigma}_\delta$ 和 $\hat{\Sigma}_{W^*}$, 则有

$$\hat{\Sigma}_\delta = \frac{1}{2m}\sum_{l=1}^{m}[W_{l1}^* - W_{l2}^*][W_{l1}^* - W_{il2}^*]^{\mathrm{T}}$$

和

$$\hat{\Sigma}_{W^*} = \frac{1}{4m}\sum_{l=1}^{m}\left\{[W_{l1}^* + W_{l2}^*][W_{l1}^* + W_{l2}^*]^{\mathrm{T}} + [W_{l1}^* - W_{l2}^*][W_{l1}^* - W_{l2}^*]^{\mathrm{T}}\right\}.$$

再根据模型 (8.3.8), 可有 $L = \Sigma_{W^*}^{-1}\mathrm{Cov}(W^*, X) = I_p - \Sigma_{W^*}^{-1}\Sigma_\delta$. 因此, 可定义在重复测量数据下 L 的估计为

$$\hat{L}_{\mathrm{RE}} = I_p - \hat{\Sigma}_{W^*}^{-1}\hat{\Sigma}_\delta. \tag{8.3.9}$$

直接使用估计量 $\hat{L}_{\mathrm{RE}}$ 作变换 $\hat{V}_{\mathrm{RE},i} = \hat{L}_{\mathrm{RE}}^{\mathrm{T}}W_i, i = 1, \cdots, n$. 类似地, 使用最小二乘方法直接得到 b 的估计 $\breve{b}_{\mathrm{LS}}$, 定义为

$$\breve{b}_{\mathrm{LS}} = \hat{\Sigma}_{V\diamond}^{-1}\frac{1}{n}\sum_{i=1}^{n}Y_i\left(\hat{V}_{\mathrm{RE},i} - \overline{\hat{V}}_{\mathrm{RE}}\right), \tag{8.3.10}$$

其中 $\hat{\Sigma}_{V\diamond} = \frac{1}{n}\sum_{i=1}^{n}\left(\hat{V}_{\mathrm{RE},i} - \overline{\hat{V}}_{\mathrm{RE}}\right)\left(\hat{V}_{\mathrm{RE},i} - \overline{\hat{V}}_{\mathrm{RE}}\right)^{\mathrm{T}}, \overline{\hat{V}}_{\mathrm{RE}} = \frac{1}{n}\sum_{i=1}^{n}\hat{V}_{\mathrm{RE},i}$.

8.3.2　切片逆回归方法

8.2.1 节已经介绍了 Li (1991) 提出的切片逆回归 (SIR) 方法. 若 L 已知时, 定义 $V = L^{\mathrm{T}}W$, 本节将考虑在测量误差模型 (8.3.1) 下, Carroll 和 Li (1992) 基于变换后的变量 (V,Y) 来给出测量误差模型下的 SIR 估计. 此时, 若假设 $E(W)=0$, $E(X)=0$, 8.2.1 节中的降维矩阵 Λ_{IR} 就变成了

$$\Lambda_{\mathrm{IR}} = \Sigma_V^{-1} E_Y \left[E(V|Y)E(V^{\mathrm{T}}|Y)\right] \Sigma_V^{-1},$$

其中 $\Sigma_V = \mathrm{Cov}(V,V) = L^{\mathrm{T}}\mathrm{Cov}(W,W)L$. 若 L 未知的时候, 可根据 8.3.1 节讨论的核实数据和重复测量误差数据情形下给出 L 的估计, 详细的讨论见 Carroll 和 Li (1992).

下面首先给出结构维数 $k=1$ 情形时, SIR 估计在 L 已知时的具体算法. 假设 $\{(W_i,Y_i),\, i=1,\cdots,n\}$ 是来自于模型 (8.3.1) 的一组 i.i.d. 的随机样本.

步骤 1　首先将数据 $\{Y_1,\cdots,Y_n\}$ 按照其取值范围分为 H 个区间: $C_{n1},\cdots,C_{nH}$. 对每个选定的区间 C_{nj}, 记 n_j 为区间 C_{nj} 内的样本个数, 并定义 $\hat{p}_j = \dfrac{n_j}{n}$, $j=1,\cdots,H$.

步骤 2　对于第 j 个区间 C_{nj} 的数据集 $\{(Y_i, V_i = L^{\mathrm{T}}W_i)$, 其中 $i=1,\cdots,n$, 并有 $Y_i \in C_{nj}\}$, 计算组内样本均值 $\bar{V}_j = \dfrac{1}{n_j}\sum\limits_{i:Y_i\in C_{nj}} V_i$, 并计算组间合并方差:

$$\hat{\Sigma}_H = \sum_{j=1}^{H} \hat{p}_j(\bar{V}_j - \bar{V})(\bar{V}_j - \bar{V})^{\mathrm{T}}, \quad \bar{V} = \frac{1}{n}\sum_{s=1}^{n} V_s.$$

步骤 3　定义 $\hat{\Sigma}_V = \dfrac{1}{n}\sum\limits_{i=1}^{n}\left(V_i - \bar{V}\right)\left(V_i - \bar{V}\right)^{\mathrm{T}}$. 若结构维数 $k=1$, 结合组间合并方差 $\hat{\Sigma}_H$ 计算特征向量 $\hat{b}_{\mathrm{SIR}}$ 为

$$\hat{\Sigma}_V^{-1}\hat{\Sigma}_H\hat{\Sigma}_V^{-1}\hat{b}_{\mathrm{SIR}} = \hat{\lambda}_{\max}\hat{b}_{\mathrm{SIR}}, \tag{8.3.11}$$

其中 $\hat{\lambda}_{\max}$ 是最大特征值. Carroll 和 Li (1992) 证明了 $\hat{b}_{\mathrm{SIR}}$ 以 $\sqrt{n}$ 的速度收敛到 $c_0\beta_1$, 其中 c_0 是一个正的常数.

若结构维数 $k>1$, 此时需进行如下的特征值分解

$$\hat{\Sigma}_V^{-1}\hat{\Sigma}_H\hat{\Sigma}_V^{-1}\hat{b}_s = \hat{\lambda}_s\hat{b}_s, \quad \hat{\lambda}_1 \geqslant \hat{\lambda}_2 \geqslant \cdots \geqslant \hat{\lambda}_k > 0.$$

Carroll 和 Li (1992) 证明了由特征向量 $\{\hat{b}_s, s=1,\cdots,k\}$ 所张成的空间 $\mathcal{S}(\hat{b}_1,\hat{b}_2,\cdots,\hat{b}_k)$ 是 $\mathcal{S}(B)$ 的相合估计.

8.3.3 累积分布降维方法

Zhang 等 (2014c) 把 Zhu 等 (2010) 提出的累积分布降维方法推广到了测量误差模型 (8.3.1), 进行了详细的研究. 不失一般性, 假设 $E(X)=0$ 和 $E(W)=0$, 在条件独立性假设 (8.1.1) 和线性条件 (8.2.1) 成立时, 结合 $L=\Sigma_W^{-1}\mathrm{Cov}(W,X)$, $\Sigma_V=L^{\mathrm{T}}\Sigma_W L$ 和 $\mathrm{Cov}(W,X)=\Gamma\Sigma_X$, 可得

$$\begin{aligned}\Sigma_V^{-1}E[I\{Y\leqslant y\}V]&=\Sigma_V^{-1}L^{\mathrm{T}}\Gamma\left\{E\Big[I\{Y\leqslant y\}E[X|(Y,B^{\mathrm{T}}X)]\Big]\right\}\\&=\Sigma_V^{-1}L^{\mathrm{T}}\Gamma\left\{E\Big[I\{Y\leqslant y\}E[X|B^{\mathrm{T}}X]\Big]\right\}\\&=\Sigma_V^{-1}L^{\mathrm{T}}\Gamma\Sigma_X B(B^{\mathrm{T}}\Sigma_X B)^{-1}E\left[I\{Y\leqslant y\}B^{\mathrm{T}}X\right],\\&=B(B^{\mathrm{T}}\Sigma_X B)^{-1}E\left[I\{Y\leqslant y\}B^{\mathrm{T}}X\right]\subseteq\mathcal{S}(B).\end{aligned}\tag{8.3.12}$$

式 (8.3.12) 表明 $\Sigma_V^{-1}E[I\{Y\leqslant y\}V]$ 属于空间 $\mathcal{S}(B)$. 若结构维数 $k=1$(即 $B=\beta_1$), 则式 (8.3.12) 可简化为

$$\Sigma_V^{-1}E[I\{Y\leqslant y\}V]=c_0(y)\beta_1,\tag{8.3.13}$$

其中 $c_0(y)=\dfrac{E[I\{Y\leqslant y\}\beta_1^{\mathrm{T}}X]}{\beta_1^{\mathrm{T}}\Sigma_X\beta_1}$. 式 (8.3.13) 表明 $\Sigma_V^{-1}E[I\{Y\leqslant y\}V]$ 与 β_1 成比例.

由式 (8.3.12), 可以定义如下的降维矩阵:

$$\Lambda=\Sigma_V^{-1}E\left\{E[I\{Y\leqslant Y^*\}V]E[I\{Y\leqslant Y^*\}V^{\mathrm{T}}]\right\}\Sigma_V^{-1},\tag{8.3.14}$$

其中 Y^* 是与 Y 独立且同分布的随机变量. 不难看出, 若结构维数 $k=1$, 在式 (8.3.13) 的基础上可以得到 $\Lambda=E[c_0^2(Y^*)]\beta_1\beta_1^{\mathrm{T}}$. 若 ζ 是矩阵 Λ 的最大特征值对应的特征向量, 那么有 $\zeta=\pm\beta_1$.

假设 $\{(W_i,Y_i),i=1,\cdots,n\}$ 是来自于模型 (8.3.1) 的一组 i.i.d. 的随机样本. 在 L 已知的情况下, 使用数据 $\{(Y_i,V_i=L^{\mathrm{T}}W_i),i=1,\cdots,n\}$, 定义 Λ 如下的估计量:

$$\hat\Lambda=\hat\Sigma_V^{-1}\frac{1}{n}\sum_{s=1}^{n}\left[\frac{1}{n}\sum_{i=1}^{n}\sum_{j=1}^{n}I\{Y_i\leqslant Y_s\}I\{Y_j\leqslant Y_s\}(V_i-\bar V)(V_j-\bar V)^{\mathrm{T}}\right]\hat\Sigma_V^{-1},$$

$$\hat\Sigma_V=\frac{1}{n}\sum_{i=1}^{n}(V_i-\bar V)(V_i-\bar V)^{\mathrm{T}},\quad \bar V=\frac{1}{n}\sum_{i=1}^{n}V_i.$$

若降维矩阵的结构维数等于 k, 则对估计的降维矩阵 $\hat\Lambda$ 进行特征值分解, 可得到特征向量 $\hat b_1,\hat b_2,\cdots,\hat b_k$, 即

$$\hat\Lambda\hat b_s=\hat\lambda_s\hat b_s,\quad \hat\lambda_1\geqslant\hat\lambda_2\geqslant\cdots\geqslant\hat\lambda_k>0.\tag{8.3.15}$$

当矩阵 L 未知时, 读者可按照 8.3.1 节中提出的在核实数据情形和重复测量数据情形来给出矩阵 L 相应的估计. Zhang 等 (2014c) 给出了矩阵 $\hat{\Lambda}$ 和特征向量 $\hat{b}_1, \hat{b}_2, \cdots, \hat{b}_k$ 在协变量 X 的维数 p 趋于无穷下的理论结果.

为了给出估计量的渐近结果, 列出了定理所需要的正则条件.

(C8.3.1) 当 $p \to \infty$ 时, $\max\limits_{1\leqslant i\leqslant p} E|X_i|^8 < \infty$ 和 $\max\limits_{1\leqslant i\leqslant p} E|\delta_i|^8 < \infty$ 成立.

(C8.3.2) 记 $\lambda_{\max,p}(LL^{\mathrm{T}})$ 是矩阵 LL^{T} 的最大特征值, $\lambda_{\max,p}(\Sigma_W)$ 是矩阵 Σ_W 的最大特征值. 那么存在正常数 C, 当 $p \to \infty$ 时, $\sup\limits_p \lambda_{\max,p}(LL^{\mathrm{T}}) < C$ 和 $\sup\limits_p \lambda_{\max,p}(\Sigma_W) < C$ 成立.

(C8.3.3) $\lambda_{\max,p}(\Sigma_X)$ 是矩阵 Σ_X 的最大特征值, $\lambda_{\max,p}(\Sigma_U)$ 是矩阵 Σ_U 的最大特征值, $\lambda_{\min,p}(\Sigma_X)$ 是矩阵 Σ_X 的最小特征值, $\lambda_{\min,p}(\Sigma_U)$ 是矩阵 Σ_U 的最小特征值, 那么存在正常数 c_1 和 c_2, 使得对任意的正整数 $p > 0$, 都有下面的不等式成立:

$$0 < c_1 < \lambda_{\min,p}(\Sigma_X) \leqslant \lambda_{\max,p}(\Sigma_X) < c_2 < \infty,$$

$$0 < c_1 < \lambda_{\min,p}(\Sigma_U) \leqslant \lambda_{\max,p}(\Sigma_U) < c_2 < \infty.$$

定理 8.3.2　*假设正则条件(C8.3.1)～ 条件(C8.3.3)成立, 则有下面的结论成立.*

(1) 若协变量 X 的维数 p 满足当 $n \to \infty$ 时, $p\log n/\sqrt{n} \to 0$, 则有

$$\frac{\sqrt{n}}{p\log n}\|\hat{\Lambda} - \Lambda\|_{\mathrm{F}} = o(1), \qquad \text{a.s.},$$

其中 $\|A\|_{\mathrm{F}}$ 表示矩阵 A 的 Frobenius 范数.

(2) 令 e 是一个长度为 p 的单位列向量, $e \in \mathcal{S}^{\perp}(B)$, 其中 $\mathcal{S}^{\perp}(B)$ 是 $\mathcal{S}(B)$ 的正交补空间. 由式 (8.3.15) 得到的特征向量 $\hat{b}_s$ 满足: 当 $p^3/n \to 0$ 成立时, $e^{\mathrm{T}}\hat{b}_s$ 以 $\sqrt{n}$ 的速度收敛到均值为 0 的正态分布, $s = 1, \cdots, k$.

定理 8.3.2 的结论 (1) 指出, 当协变量 X 的维数 p 以 $o(\sqrt{n}/\log n)$ 的速度趋于无穷时, $\hat{\Lambda}$ 几乎处处收敛到累积分布降维矩阵 Λ, 该结论为累积分布降维方法处理高维数据提供了一定的理论基础. 定理 8.3.2 的结论 (2) 指出, 当协变量 X 的维数 p 以 $o\left(n^{1/3}\right)$ 的速度趋于无穷时, 降维矩阵 $\hat{\Lambda}$ 的特征向量 $\hat{b}_s$ 是渐近正态的, $s = 1, \cdots, k$.

8.3.4　Hessian 主方向方法

Lue (2004) 提出了在测量误差结构 (8.3.1) 下的 Hessian 主方向 (PHD) 方法. 不失一般性, 假设 $E(X) = 0$ 和 $E(W) = 0$. 正如 8.2.5 节介绍的 PHD 方法, 当 X 服从多元正态分布时, Li (1992) 利用 Stein 引理 (Stein, 1981) 证明了降维矩阵 H_X 的非零特征向量可用于求解降维空间 $\mathcal{S}(B)$. 降维矩阵 H_X 的定义如下:

$$H_X = \Sigma_X^{-1} E\left[(Y - E(Y))XX^{\mathrm{T}}\right] \Sigma_X^{-1}.$$

在测量误差结构 (8.3.1) 下, 则有如下恒等式:

$$\begin{aligned} H_V =&: \Sigma_V^{-1} E[(Y - E(Y)) V V^{\mathrm{T}}] \Sigma_V^{-1} \\ &= \Sigma_V^{-1} L^{\mathrm{T}} E[(Y - E(Y)) W W^{\mathrm{T}}] L \Sigma_V^{-1} \\ &= \Sigma_V^{-1} L^{\mathrm{T}} \Gamma E[(Y - E(Y)) X X^{\mathrm{T}}] \Gamma L^{\mathrm{T}} \Sigma_V^{-1} \\ &= \Sigma_X^{-1} E[H(X)] \Sigma_X^{-1} = H_X. \end{aligned} \tag{8.3.16}$$

因此, 从式 (8.3.16) 不难得到, 若 $v \in \mathcal{S}^{\perp}(B)$, 则有 $v^{\mathrm{T}} H_X = 0$, 因此也就得到了 $v^{\mathrm{T}} H_V = 0$. 从而矩阵 H_V 的前 k 个非零特征值 $\lambda_s (s = 1, \cdots, k)$ 对应的特征向量 b_s 可用来求解降维空间 $\mathcal{S}(B)$, 即 b_s 满足

$$H_V b_s = \lambda_s b_s, \quad |\lambda_1| \geqslant |\lambda_2| \geqslant \cdots \geqslant |\lambda_k| > 0.$$

下面给出矩阵 H_V 在 L 已知时的估计 $\hat{H}_V$, 和估计 $\hat{H}_V$ 的特征值分解用于求解降维空间 $\mathcal{S}(B)$. 假设 $\{(W_i, Y_i), i = 1, \cdots, n\}$ 是来自于模型 (8.3.1) 的一组 i.i.d. 的随机样本. 在 L 已知的情况下, 使用数据 $\{(Y_i, V_i = L^{\mathrm{T}} W_i), i = 1, \cdots, n\}$ 定义如下估计:

$$\begin{aligned} \hat{H}_V &= \hat{\Sigma}_V^{-1} \left\{ L^{\mathrm{T}} \frac{1}{n} \sum_{i=1}^{n} \left[(Y_i - \bar{Y})(V_i - \bar{V})(V_i - \bar{V})^{\mathrm{T}} \right] L \right\} \hat{\Sigma}_V^{-1}, \\ \hat{\Sigma}_V &= \frac{1}{n} \sum_{i=1}^{n} (V_i - \bar{V})(V_i - \bar{V})^{\mathrm{T}}, \qquad \bar{V} = \frac{1}{n} \sum_{i=1}^{n} V_i, \\ \hat{H}_V \hat{b}_s &= \hat{\lambda}_s \hat{b}_s, \qquad |\hat{\lambda}_1| \geqslant |\hat{\lambda}_2| \geqslant \cdots \geqslant |\hat{\lambda}_k| > 0. \end{aligned}$$

因此, 由 k 个特征向量 $\hat{b}_1, \hat{b}_2, \cdots, \hat{b}_k$ 所张成的空间 $\mathcal{S}(\hat{b}_1, \hat{b}_2, \cdots, \hat{b}_k)$ 可用来估计 $\mathcal{S}(B)$. 当矩阵 L 未知时, 读者可按照 8.3.1 节中提出的在核实数据情形和重复测量数据情形来给出矩阵 L 相应的估计.

8.3.5 方向回归方法

8.2.4 节中介绍了 Li 和 Wang (2007) 提出了方向回归方法, 在线性条件 (8.2.1) 和常数方差条件 (8.2.7) 成立的情况下, Li 和 Wang (2007) 证明了

$$V(Y_1, Y_2) = \Sigma_X^{-1} E\left[(X_1 - X_2)(X_1 - X_2)^{\mathrm{T}} \middle| Y_1, Y_2 \right] \subseteq \mathcal{S}(B), \tag{8.3.17}$$

其中 (X_1, Y_1) 和 (X_2, Y_2) 是与 (X, Y) 独立同分布的随机变量. 方向回归降维方法的降维矩阵 Λ_{DR}, 定义如下:

$$\Lambda_{\mathrm{DR}} = \Sigma_X^{-1} E_{Y_1, Y_2} \left[(2\Sigma_X - V(Y_1, Y_2))(2\Sigma_X - V(Y_1, Y_2))^{\mathrm{T}} \right] \Sigma_X^{-1}.$$

降维矩阵 $\Lambda_{\rm DR}$ 的前 k 个非零最大特征值所对应的特征向量可用来求解降维空间 $\mathcal{S}(B)$. 基于此想法, Zhu 等 (2010) 拓展了方向回归的方法, 提出了基于方向累积分布降维方法, 定义

$$\begin{aligned}N(y_1,y_2)=&\Sigma_X^{-1}E\Big[(X_1-X_2)(X_1-X_2)^{\rm T}I\{Y_1\leqslant y_1\}I\{Y_2\leqslant y_2\}\Big]\\&-2F(y_1)F(y_2)I_p.\end{aligned}\tag{8.3.18}$$

Zhu 等 (2010) 证明了 $N(y_1,y_2)\subseteq\mathcal{S}(B)$. Zhang 等 (2014c) 把 Zhu 等 (2010) 提出的基于方向累积分布降维方法推广到本节考虑的测量误差模型, 定义函数 $R(y_1,y_2)$ 如下

$$\begin{aligned}R(y_1,y_2)=&\Sigma_V^{-1}E\Big[(V_1-V_2)(V_1-V_2)^{\rm T}I\{Y_1\leqslant y_1\}I\{Y_2\leqslant y_2\}\Big]\\&-2F(y_1)F(y_2)I_p,\end{aligned}$$

其中 $V_1=L^{\rm T}W_1$, $V_2=L^{\rm T}W_2$, 并且 (V_1,Y_1) 和 (V_2,Y_2) 是与 $(L^{\rm T}W,Y)$ 独立同分布的随机变量. 则可定义如下的降维矩阵

$$M_*=E[R(Y_1,Y_2)R^{\rm T}(Y_1,Y_2)].$$

若降维空间 $\mathcal{S}(B)$ 的结构维数为 k, 可对降维矩阵 M_* 进行如下特征值分解, 即

$$M_*\eta_s=\lambda_s\eta_s,\quad \lambda_1\geqslant\lambda_2\geqslant\cdots\geqslant\lambda_k>0,$$

其中 η_s 是第 s 个非零特征值 λ_s 所对应的特征向量. Zhang 等 (2014c) 指出在线性条件 (8.2.1) 和常数方差条件 (8.2.7) 均成立时, $\mathcal{S}(\eta_1,\cdots,\eta_k)\subseteq\mathcal{S}(B)$, 即由前 k 个特征向量 $\eta_1,\cdots,\eta_k$ 所张成的列向量空间可用于求解降维空间 $\mathcal{S}(B)$.

假设 $\{(W_i,Y_i),i=1,\cdots,n\}$ 是来自于模型 (8.3.1) 的一组 i.i.d.的随机样本. 下面使用数据 $\{(Y_i,V_i=L^{\rm T}W_i),i=1,\cdots,n\}$ 分别给出降维矩阵 M_* 在 L 已知时的估计 $\hat{M}_*$, 和降维矩阵估计 $\hat{M}_*$ 的特征值分解. 降维矩阵 M_* 在 L 已知时的估计 $\hat{M}_*$ 定义如下:

$$\hat{M}_*=\frac{1}{n^2}\sum_{i=1}^n\sum_{j=1}^n\hat{R}(Y_i,Y_j)\hat{R}(Y_i,Y_j),$$

其中

$$\begin{aligned}\hat{R}(Y_i,Y_j)=&\hat{\Sigma}_V^{-1}\frac{1}{n^2}\sum_{s=1}^n\sum_{l=1}^n\left[L^{\rm T}(W_s-W_l)(W_s-W_l)^{\rm T}L\right]I\{Y_s\leqslant Y_i\}I\{Y_l\leqslant Y_j\}\\&-2\left(\frac{1}{n}\sum_{s=1}^nI\{Y_s\leqslant Y_i\}\right)\left(\frac{1}{n}\sum_{l=1}^nI\{Y_l\leqslant Y_j\}\right)I_p,\end{aligned}$$

$$\hat{\Sigma}_V=\frac{1}{n}\sum_{i=1}^n(V_i-\bar{V})(V_i-\bar{V})^{\rm T},\quad \bar{V}=\frac{1}{n}\sum_{i=1}^nV_i.$$

获得降维矩阵的估计 $\hat{M}_*$ 后, 可对估计的降维矩阵 $\hat{M}_*$ 进行如下的特征值分解:

$$\hat{M}_*\hat{b}_s = \hat{\lambda}_s\hat{b}_s, \quad \hat{\lambda}_1 \geqslant \hat{\lambda}_2 \geqslant \cdots \geqslant \hat{\lambda}_k > 0.$$

这时, 由 k 个特征向量 $\hat{b}_1, \hat{b}_2, \cdots, \hat{b}_k$ 所张成的空间 $\mathcal{S}(\hat{b}_1, \hat{b}_2, \cdots, \hat{b}_k)$ 可用来估计 $\mathcal{S}(B)$. 同样当矩阵 L 未知时, 读者可按照 8.3.1 节中提出的在核实数据情形和重复测量数据情形来给出矩阵 L 相应的估计.

8.3.6 等高线回归方法

8.2.3 节已经介绍了 Li 等 (2005) 提出的等高线回归方法, 在条件方差不等式 (8.2.10) 成立的情况下, 等高线回归降维矩阵可定义为

$$\Lambda_{\mathrm{CR}}^*(c) = \Sigma_X^{-1/2} E\left[(X_1 - X_2)(X_1 - X_2)^{\mathrm{T}} I\{|Y_1 - Y_2| \leqslant c\}\right] \Sigma_X^{-1/2},$$

其中 $\Sigma_X^{-1/2}$ 是对称矩阵且有 $\Sigma_X^{-1/2}\Sigma_X^{-1/2} = \Sigma_X^{-1}$. 若降维空间 $\mathcal{S}(B)$ 的结构维数为 k 时, 那么由矩阵 $\Lambda_{\mathrm{CR}}^*(c)$ 的前 k 个最小非零特征值所对应的特征向量所张成的线性空间可用于求解降维空间 $\mathcal{S}(B)$.

Li 和 Yin (2007) 提出了测量误差模型 (8.3.1) 下的不变准则, 若矩阵 L 已知, 在一些正则条件下, 可以证得

$$Y \perp\!\!\!\perp X | B^{\mathrm{T}}X \quad \text{等价于} \quad Y \perp\!\!\!\perp L^{\mathrm{T}}W | B^{\mathrm{T}}(L^{\mathrm{T}}W). \tag{8.3.19}$$

式 (8.3.19) 指出, 即使协变量 X 不能准确测量, 仍然可使用观测变量 $(V = L^{\mathrm{T}}W, Y)$ 通过一些降维方法对降维空间 $\mathcal{S}(B)$ 进行估计. 为此, Li 和 Yin (2007) 将不变准则的思想用于等高线回归方法研究测量误差模型.

假设 $\{(W_i, Y_i), i = 1, \cdots, n\}$ 是来自于模型 (8.3.1) 的一组 i.i.d. 的随机样本. 若 L 已知, 定义 $V_i = L^{\mathrm{T}}W_i$, $i = 1, \cdots, n$. 由 8.2.3 节提出的降维矩阵 $\Lambda_{\mathrm{CR}}^*(c)$ 的估计方法, 利用不变准则 (8.3.19) 可直接定义如下降维矩阵的估计:

$$\hat{\Lambda}_{\mathrm{CR}}^*(c) = \hat{\Sigma}_V^{-1/2}\hat{H}_n(c)\hat{\Sigma}_V^{-1/2},$$

其中

$$\hat{H}_n(c) = \frac{2}{n(n-1)} \sum_{1 \leqslant i < j \leqslant n}^{n} (V_i - V_j)(V_i - V_j)^{\mathrm{T}} I\{|Y_i - Y_j| \leqslant c\},$$

$$\hat{\Sigma}_V = \frac{1}{n}\sum_{i=1}^{n}\left(V_i - \bar{V}\right)\left(V_i - \bar{V}\right)^{\mathrm{T}}, \quad \bar{V} = \frac{1}{n}\sum_{i=1}^{n} V_i,$$

其中 $\hat{\Sigma}_V^{-1/2}$ 是对称矩阵且满足 $\hat{\Sigma}_V^{-1/2}\hat{\Sigma}_V^{-1/2} = \hat{\Sigma}_V^{-1}$. 有了估计的降维矩阵 $\hat{\Lambda}_{\mathrm{CR}}^*(c)$ 以后, 对降维矩阵 $\hat{\Lambda}_{\mathrm{CR}}^*(c)$ 进行特征值分解, 即

$$\hat{\Lambda}^*_{\rm CR}(c)\hat{\eta}_s=\hat{\lambda}_s\hat{\eta}_s,\quad \hat{\lambda}_1\geqslant\hat{\lambda}_2\geqslant\cdots\geqslant\hat{\lambda}_{p-1}\geqslant\hat{\lambda}_p.$$

对前 k 个最小特征值 $\hat{\lambda}_{p-k+1},\hat{\lambda}_{p-k+2},\cdots,\hat{\lambda}_p$ 所对应的特征向量 $\hat{\eta}_{p-k+1}$, $\hat{\eta}_{p-k+2},\cdots,\hat{\eta}_p$ 进行线性变换: $\hat{\Sigma}_V^{1/2}\hat{\eta}_{p-k+1},\hat{\Sigma}_V^{1/2}\hat{\eta}_{p-k+2},\cdots,\hat{\Sigma}_V^{1/2}\hat{\eta}_p$, 其中 $\hat{\Sigma}_V^{1/2}$ 是对称矩阵且满足 $\hat{\Sigma}_V^{1/2}\hat{\Sigma}_V^{1/2}=\hat{\Sigma}_V$. Li 和 Yin (2007) 使用变换之后的特征向量所张成的空间 $\mathcal{S}(\hat{\Sigma}_V^{1/2}\ \hat{\eta}_{p-k+1},\hat{\Sigma}_V^{1/2}\hat{\eta}_{p-k+2},\cdots,\hat{\Sigma}_V^{1/2}\hat{\eta}_p)$ 可作为降维空间 $\mathcal{S}(B)$ 的估计.

8.3.7 模拟研究

本节通过一些模拟研究对切片逆回归 (SIR) 方法, Hessian 主方向 (PHD) 方法, 等高线回归 (CR) 方法和累积分布降维 (CU) 方法进行比较, 其中 SIR 方法的切片数取 5 和 10. 为了比较各种方法的优劣, 采用迹相关系数准则 (Férre, 1998), 定义如下:

$$R^2(k)=\frac{1}{k}\mathrm{tr}\left(\Pi_B\Pi_{\widehat{B}}\right),\tag{8.3.20}$$

其中 k 是结构维数, 矩阵 $\hat{B}$ 是维数为 $p\times k$ 的矩阵. 对不同的降维方法而言, 矩阵 $\hat{B}$ 的列向量是由 SIR 方法、PHD 方法、CR 方法和 CU 方法的前 k 个非零特征值所对应的特征向量组成的. Π_B 和 $\Pi_{\widehat{B}}$ 分别是矩阵 B 和 $\hat{B}$ 的投影矩阵. 迹相关准则 $R^2(k)$ 的值越靠近 1, 表明该降维方法的估计效果越有效. 模拟结果报告了重复 200 次模拟后迹相关系数 $R^2(k)$ 的平均均值和标准差.

本例将协变量 X 的维数设定为 $p=19,29,39,49$, 用于考察估计方法在协变量维数较高时的估计效果. 考虑下面 6 个模型:

$$Y=X_1-X_2+X_3+4\times\varepsilon,\tag{8.3.21}$$

$$Y=\exp(X_1-X_2+4\times\varepsilon),\tag{8.3.22}$$

$$Y=\sin(X_1-X_2+1.5\times\varepsilon),\tag{8.3.23}$$

$$Y=\log(|X_1+1|)+\varepsilon,\tag{8.3.24}$$

$$Y=X_1(1+X_2)+0.5\times\varepsilon,\tag{8.3.25}$$

$$Y=X_2/[0.5+(1.5+X_1)^2]+\varepsilon.\tag{8.3.26}$$

上面 6 个模型中, $X=(X_1,X_2,\cdots,X_p)^{\rm T}$ 和误差项 ε 独立. 在模型 (8.3.21) 中, 矩阵 B 取为 $B=\beta_1$, 其中 $\beta_1=(1,-1,1,0,0,\cdots,0)^{\rm T}\in\mathbb{R}^{p\times1}$. 对模型 (8.3.22) 和模型 (8.3.23), $B=\beta_1$, 其中 $\beta_1=(1,-1,0,0,\cdots,0)^{\rm T}\in\mathbb{R}^{p\times1}$. 对模型 (8.3.24), $B=\beta_1$, 其中 $\beta_1=(1,0,0,\cdots,0)^{\rm T}\in\mathbb{R}^{p\times1}$. 对模型 (8.3.25) 和模型 (8.3.26), $B=(\beta_1,\beta_2)$, $\beta_1=(1,0,0,\cdots,0)^{\rm T}\in\mathbb{R}^{p\times1},\beta_2=(0,1,0,0,\cdots,0)^{\rm T}\in\mathbb{R}^{p\times1}$.

情形 1 L 已知的情形. 协变量 X_{ij} 独立同分布取自于 $t(6)$ 分布, $i=1,\cdots,n$, $j=1,\cdots,p$. 模型误差 ε 取自均匀分布 $(-0.2,0.2)$. 将测量误差模型 (8.3.1) 中的 Γ 取为对角线元素为 1, 其余位置全为 0.5 的 $p\times p$ 对称矩阵. 误差 δ_i 服从正态分布 $N_p(0,0.3^2\times I_p)$, $i=1,\cdots,n$. 此时, $L=\Gamma^{\mathrm{T}}\left(\Gamma\Gamma^{\mathrm{T}}+\dfrac{0.3^2}{\sigma^2_{t(6)}}\times I_p\right)^{-1}$, 其中 $\sigma^2_{t(6)}$ 是 $t(6)$ 分布的方差. 表 8.3.1 报告了在矩阵 L 已知情形下, 重复 200 次模拟后迹相关系数 $R^2(k)$ 的平均均值和标准差.

表 8.3.1 情形 1: 矩阵 L 已知情形下, 迹相关系数 $R^2(k)$ 的模拟结果

(n,p)	模型		SIR		PHD	CR	CU
			5	10			
(100,19)	模型(8.3.21)	均值	0.8613	0.8733	0.1547	0.8431	0.8598
		标准差	0.0436	0.0431	0.1248	0.0467	0.0407
	模型(8.3.22)	均值	0.8560	0.8714	0.2936	0.8441	0.8545
		标准差	0.0508	0.0445	0.1652	0.0494	0.0487
	模型(8.3.23)	均值	0.7489	0.7502	0.0356	0.7641	0.7745
		标准差	0.1459	0.1422	0.0497	0.1201	0.1112
	模型(8.3.24)	均值	0.6271	0.6171	0.0872	0.6277	0.6698
		标准差	0.2122	0.2306	0.1684	0.1610	0.1430
	模型(8.3.25)	均值	0.4557	0.4372	0.3666	0.5213	0.4967
		标准差	0.1191	0.1157	0.1205	0.1007	0.1101
	模型(8.3.26)	均值	0.5955	0.5807	0.2253	0.4955	0.6738
		标准差	0.1028	0.1025	0.1066	0.1201	0.0894
(225,29)	模型(8.3.21)	均值	0.9046	0.9009	0.1809	0.8742	0.8990
		标准差	0.0266	0.0238	0.1369	0.0301	0.0243
	模型(8.3.22)	均值	0.9065	0.9128	0.2634	0.8952	0.9058
		标准差	0.0310	0.0247	0.1315	0.0288	0.0227
	模型(8.3.23)	均值	0.8348	0.8444	0.0305	0.8477	0.8563
		标准差	0.0682	0.0602	0.0505	0.0620	0.0580
	模型(8.3.24)	均值	0.7562	0.7591	0.0918	0.7445	0.7813
		标准差	0.1040	0.1005	0.2297	0.0987	0.0702
	模型(8.3.25)	均值	0.5422	0.5412	0.4329	0.6524	0.6307
		标准差	0.1118	0.1113	0.1185	0.0743	0.0811
	模型(8.3.26)	均值	0.7212	0.7215	0.2303	0.5122	0.7588
		标准差	0.0642	0.0622	0.0948	0.0899	0.0491
(400,39)	模型(8.3.21)	均值	0.9324	0.9244	0.1646	0.9132	0.9206
		标准差	0.0166	0.0143	0.1178	0.0201	0.0138
	模型(8.3.22)	均值	0.9313	0.9317	0.2826	0.9195	0.9295
		标准差	0.0165	0.0161	0.1184	0.0181	0.0143
	模型(8.3.23)	均值	0.8678	0.8821	0.0223	0.8700	0.8859
		标准差	0.0427	0.0367	0.0427	0.0421	0.0347

续表

(n,p)	模型		SIR 5	SIR 10	PHD	CR	CU
(400,39)	模型(8.3.24)	均值	0.8240	0.8239	0.0893	0.7723	0.8175
		标准差	0.0610	0.0650	0.2324	0.0711	0.0506
	模型(8.3.25)	均值	0.6371	0.6459	0.4983	0.7134	0.6929
		标准差	0.0715	0.0912	0.1037	0.0561	0.0616
	模型(8.3.26)	均值	0.7752	0.7979	0.2661	0.5799	0.8017
		标准差	0.0398	0.0463	0.0916	0.0654	0.0366
(625,49)	模型(8.3.21)	均值	0.9464	0.9551	0.1745	0.9326	0.9466
		标准差	0.0096	0.0104	0.1153	0.0132	0.0098
	模型(8.3.22)	均值	0.9441	0.9439	0.3029	0.9276	0.9376
		标准差	0.0120	0.0114	0.1175	0.0141	0.0111
	模型(8.3.23)	均值	0.9003	0.9015	0.0183	0.8923	0.9046
		标准差	0.0282	0.0262	0.0289	0.0291	0.0233
	模型(8.3.24)	均值	0.8526	0.8580	0.2064	0.8188	0.8588
		标准差	0.0502	0.0495	0.3319	0.0497	0.0433
	模型(8.3.25)	均值	0.6969	0.7228	0.5558	0.7577	0.7384
		标准差	0.0578	0.0567	0.0945	0.0366	0.0494
	模型(8.3.26)	均值	0.8124	0.8342	0.2917	0.6066	0.8366
		标准差	0.0369	0.0323	0.0950	0.0539	0.0299

从表 8.3.1 中的模拟结果可以看出：

(1) 对模型 (8.3.21) 和模型 (8.3.22), 切片逆回归方法的估计效果非常好, 这个结论已经被一些文献中的模拟结果所证实;

(2) 对模型 (8.3.23)~模型 (8.3.26), 累积分布降维方法的估计效果要远远好于切片逆回归方法. 但是值得一提的是, 切片逆回归方法中不同的切片数会得到不同的模拟结果, 但是累积分布降维方法不需要对切片数进行选取, 效果也很稳定. 在实际操作时, 累积分布降维方法更容易实施;

(3) 对于所有的模型, Hessian 主方向方法的模拟效果都不是很好, 即使样本量增加到 625. 对模型 (8.3.25), 等高线回归方法要好于其他几种方法;

(4) 随着协变量 X 维数 p 的增加, 累积分布降维方法对模型 (8.3.21)~模型 (8.3.26) 的模拟结果总体表现要好于其他几种方法.

情形 2　L 未知时使用核实数据估计 L 的情形. 此时变量 $(X,\delta,\varepsilon,\Gamma)$ 的随机产生方式和情形 1 相同. 同时, 独立产生另外一组核实数据, 该核实数据的产生方法和情形 1 也完全相同, 核实数据的样本量选取为 $m=[n/2]$, 其中 $[\cdot]$ 是取整函数. 首先由式 (8.3.6) 和这组核实数据, 可以得到估计 $\hat{L}_{\mathrm{LS}}$, 然后使用数据 $\{(Y_i,\hat{V}_i=\hat{L}_{\mathrm{LS}}^{\mathrm{T}}W_i),i=1,\cdots,n\}$ 用来估计降维空间 $\mathcal{S}(B)$. 表 8.3.2 报告了重复 200 次模拟后迹相关系数 $R^2(k)$ 的平均均值和标准差.

表 8.3.2 情形 2: 使用核实数据情形下, 迹相关系数 $R^2(k)$ 的模拟结果

(n,p,m)	模型		SIR 5	SIR 10	PHD	CR	CU
(100,19,50)	模型(8.3.21)	均值	0.6989	0.7253	0.1328	0.6990	0.7190
		标准差	0.1026	0.0869	0.1176	0.0993	0.0967
	模型(8.3.22)	均值	0.7145	0.7062	0.2096	0.6823	0.7077
		标准差	0.0954	0.0988	0.1533	0.1290	0.1023
	模型(8.3.23)	均值	0.6159	0.5882	0.0445	0.6123	0.6313
		标准差	0.1565	0.1515	0.0621	0.1411	0.1314
	模型(8.3.24)	均值	0.5264	0.5145	0.0835	0.4923	0.5715
		标准差	0.1668	0.2109	0.1702	0.1689	0.1473
	模型(8.3.25)	均值	0.3904	0.3726	0.3368	0.4756	0.4406
		标准差	0.1229	0.1163	0.1189	0.0923	0.1028
	模型(8.3.26)	均值	0.5169	0.5322	0.2217	0.4239	0.5811
		标准差	0.1051	0.1073	0.1028	0.1106	0.0855
(225,29,112)	模型(8.3.21)	均值	0.8409	0.8464	0.1513	0.8298	0.8355
		标准差	0.0389	0.0423	0.1191	0.0513	0.0487
	模型(8.3.22)	均值	0.8277	0.8356	0.2422	0.8131	0.8325
		标准差	0.0457	0.0394	0.1242	0.0617	0.0511
	模型(8.3.23)	均值	0.7513	0.7578	0.0278	0.7607	0.7703
		标准差	0.0824	0.0710	0.0385	0.0734	0.0679
	模型(8.3.24)	均值	0.7167	0.7000	0.0797	0.6945	0.7275
		标准差	0.0937	0.0864	0.1814	0.0911	0.0824
	模型(8.3.25)	均值	0.5158	0.5113	0.4061	0.5756	0.5502
		标准差	0.0896	0.1005	0.1126	0.0744	0.0859
	模型(8.3.26)	均值	0.6634	0.6766	0.2383	0.5433	0.6915
		标准差	0.0723	0.0696	0.0918	0.0772	0.0646
(400,39,200)	模型(8.3.21)	均值	0.8771	0.8867	0.1316	0.8600	0.8679
		标准差	0.0295	0.0247	0.1114	0.0379	0.0279
	模型(8.3.22)	均值	0.8725	0.8620	0.2885	0.8270	0.8569
		标准差	0.0298	0.0301	0.1168	0.0334	0.0301
	模型(8.3.23)	均值	0.8275	0.8248	0.0246	0.8055	0.8324
		标准差	0.0535	0.0524	0.0335	0.0489	0.0467
	模型(8.3.24)	均值	0.7717	0.7775	0.1205	0.7461	0.7864
		标准差	0.0686	0.0664	0.2596	0.0609	0.0574
	模型(8.3.25)	均值	0.6063	0.6176	0.4650	0.6734	0.6505
		标准差	0.0825	0.0905	0.1093	0.0540	0.0628
	模型(8.3.26)	均值	0.7411	0.7530	0.2574	0.5876	0.7643
		标准差	0.0468	0.0455	0.0884	0.0563	0.0487
(625,49,312)	模型(8.3.21)	均值	0.9041	0.9049	0.1422	0.8887	0.8997
		标准差	0.0214	0.0192	0.0961	0.0275	0.0201
	模型(8.3.22)	均值	0.9017	0.9239	0.2861	0.8882	0.9178
		标准差	0.0214	0.0222	0.0995	0.0275	0.0203

续表

(n,p,m)	模型		SIR 5	SIR 10	PHD	CR	CU
(625,49,312)	模型(8.3.23)	均值	0.8600	0.8616	0.0150	0.8467	0.8632
		标准差	0.0398	0.0375	0.0191	0.0311	0.0308
	模型(8.3.24)	均值	0.8210	0.8281	0.1754	0.8089	0.8301
		标准差	0.0549	0.0453	0.3037	0.0422	0.0388
	模型(8.3.25)	均值	0.6737	0.6854	0.5149	0.7233	0.6999
		标准差	0.0601	0.0631	0.1081	0.0473	0.0538
	模型(8.3.26)	均值	0.7862	0.8024	0.2939	0.6123	0.8045
		标准差	0.0341	0.0329	0.0905	0.0454	0.0317

从表 8.3.2 中的模拟结果可以看出：切片逆回归方法对模型 (8.3.21) 和模型 (8.3.22) 的模拟效果仍旧是最好的. 对模型 (8.3.25), 等高线回归方法要好于其他几种方法; 此外, 累积分布降维方法的结果总体表现要好于其他几种方法. 同时也不难看出, 表 8.3.2 中的模拟结果没有在 L 已知情形下表 8.3.1 中的模拟结果好, 这是因为这里需要估计未知矩阵 L.

情形 3　L 未知时使用重复测量误差数据估计 L 的情形. 此时变量 $(X,\delta,\varepsilon,\Gamma)$ 的随机产生方式和情形 1 相同. 同时, 独立产生另外一组重复测量误差数据, 该数据来自于模型 $W_{ij}=X_i+\delta_{ij}, j=1,2, i=1,\cdots,m$. 协变量 X_{ij} 服从 $t(6)$ 分布, $i=1,\cdots,m, j=1,\cdots,p$. δ_{ij} 取自正态分布 $N(0,0.3^2)$, $i=1,\cdots,m, j=1,2$. 模型误差 ε 来自于区间 $(-0.2,0.2)$ 上的均匀分布. 重复测量误差数据的样本量为 $m=[n/2]$. 由 8.3.1 节式 (8.3.9), 可得到 L 的估计 $\hat{L}_{\mathrm{RE}}$, 这时使用数据 $\{(Y_i,\hat{V}_i=\hat{L}_{\mathrm{RE}}^{\mathrm{T}}W_i), i=1,\cdots,n\}$ 用来估计降维空间 $\mathcal{S}(B)$. 表 8.3.3 报告了重复 200 次模拟后迹相关系数 $R^2(k)$ 的平均均值和标准差. 从表 8.3.3 可以看出, 重复测量误差数据这种情形下的模拟结果与情形 1 和情形 2 的模拟结果很相似.

表 8.3.3　情形 3: 使用重复测量误差数据情形下, 迹相关系数 $R^2(k)$ 的模拟结果

(n,p,m)	模型		SIR 5	SIR 10	PHD	CR	CU
(100,19,50)	模型(8.3.21)	均值	0.8794	0.9021	0.1660	0.8851	0.8941
		标准差	0.0437	0.9021	0.1431	0.0458	0.0427
	模型(8.3.22)	均值	0.8776	0.8842	0.2778	0.8521	0.8755
		标准差	0.0491	0.0397	0.1536	0.0556	0.0439
	模型(8.3.23)	均值	0.7891	0.7721	0.0496	0.7700	0.7945
		标准差	0.1259	0.1473	0.0721	0.1078	0.1230
	模型(8.3.24)	均值	0.6848	0.6789	0.0806	0.6878	0.7072
		标准差	0.1749	0.2126	0.1783	0.1092	0.1316
	模型(8.3.25)	均值	0.4962	0.4400	0.4303	0.5534	0.5393
		标准差	0.1363	0.1488	0.1340	0.0899	0.1047

续表

(n,p,m)	模型		SIR		PHD	CR	CU
			5	10			
(100,19,50)	模型(8.3.26)	均值	0.6472	0.6493	0.2632	0.5023	0.7014
		标准差	0.1008	0.1137	0.1007	0.0877	0.0835
(225,29,112)	模型(8.3.21)	均值	0.9374	0.9212	0.1542	0.9167	0.9267
		标准差	0.0198	0.0143	0.1201	0.0201	0.0196
	模型(8.3.22)	均值	0.9297	0.9172	0.3091	0.8801	0.9107
		标准差	0.0221	0.0166	0.1412	0.0389	0.0245
	模型(8.3.23)	均值	0.8561	0.8521	0.0259	0.8453	0.8704
		标准差	0.0608	0.0804	0.0548	0.0412	0.0570
	模型(8.3.24)	均值	0.7979	0.8066	0.1329	0.7765	0.8101
		标准差	0.0944	0.1134	0.2719	0.0813	0.0749
	模型(8.3.25)	均值	0.6001	0.6176	0.4770	0.6856	0.6555
		标准差	0.1089	0.0998	0.1291	0.0834	0.0948
	模型(8.3.26)	均值	0.7688	0.7718	0.2784	0.6134	0.8006
		标准差	0.0590	0.0628	0.0991	0.0676	0.0484
(400,39,200)	模型(8.3.21)	均值	0.9525	0.9630	0.1742	0.9328	0.9523
		标准差	0.0126	0.0100	0.1161	0.0144	0.0128
	模型(8.3.22)	均值	0.9507	0.9433	0.3262	0.9217	0.9432
		标准差	0.0146	0.0158	0.1249	0.0134	0.0119
	模型(8.3.23)	均值	0.9040	0.8992	0.0168	0.8832	0.9074
		标准差	0.0343	0.0416	0.0327	0.0319	0.0324
	模型(8.3.24)	均值	0.8592	0.8584	0.2721	0.8133	0.8591
		标准差	0.0545	0.0531	0.3616	0.0521	0.0430
	模型(8.3.25)	均值	0.6898	0.7039	0.5599	0.7445	0.7275
		标准差	0.0702	0.0772	0.1156	0.0498	0.0618
	模型(8.3.26)	均值	0.8179	0.8244	0.3179	0.6455	0.8430
		标准差	0.0419	0.0427	0.0948	0.0445	0.0387
(625,49,312)	模型(8.3.21)	均值	0.9632	0.9598	0.1928	0.9401	0.9449
		标准差	0.0079	0.0072	0.1062	0.0085	0.0079
	模型(8.3.22)	均值	0.9595	0.9588	0.3401	0.9389	0.9489
		标准差	0.0105	0.0089	0.1145	0.0121	0.0089
	模型(8.3.23)	均值	0.9217	0.9224	0.0138	0.9004	0.9228
		标准差	0.0246	0.0236	0.0194	0.0187	0.0216
	模型(8.3.24)	均值	0.8739	0.8855	0.3370	0.8654	0.8857
		标准差	0.0414	0.0403	0.3937	0.0423	0.0368
	模型(8.3.25)	均值	0.7301	0.7554	0.6136	0.7823	0.7615
		标准差	0.0574	0.0570	0.1042	0.0396	0.0448
	模型(8.3.26)	均值	0.8455	0.8628	0.3476	0.6676	0.8652
		标准差	0.0309	0.0276	0.1070	0.0348	0.0272

8.4 协变量调整降维模型

8.4.1 模型介绍

本节考虑如下的协变量调整降维模型：

$$\begin{cases} Y \perp\!\!\!\perp X | B^{\mathrm{T}}X, \\ \widetilde{Y} = \phi(U)Y, \\ \widetilde{X}_r = \psi_r(U)X_r, \quad r = 1, \cdots, p, \end{cases} \tag{8.4.1}$$

其中 Y 是不可观测的响应变量, $X = (X_1, X_2, \cdots, X_p)^{\mathrm{T}}$ 是不可观测的连续型协变量向量, $\widetilde{Y}$ 和 $\widetilde{X}_r$ 分别是测量到的响应变量和协变量, U 是一个 1 维可观测的混淆变量, $\phi(U)$ 和 $\psi_r(U)$ 是混淆变量 U 的未知的失真函数, $r = 1, \cdots, p$. $B = (\beta_1, \beta_2, \cdots, \beta_k)$ 是未知的 $p \times k$ 列满秩矩阵, 并有 $B^{\mathrm{T}}B = I_k$. 这里的 k 是降维空间 $\mathcal{S}(B)$ 的结构维数. 为了介绍估计程序, 需引入如下一些关于模型 (8.4.1) 的两个基本假设条件：

(1) $E[\phi(U)] = 1$, $E[\psi_r(U)] = 1, r = 1, \cdots, p$;

(2) (X, U) 是相互独立的, 且 (Y, U) 也是相互独立的.

8.4.2 估计方法和渐近结果

首先介绍降维空间 $\mathcal{S}(B)$ 的估计方法. 假设线性条件 (8.2.1) 和混淆变量 U 与 (X, Y) 相互独立这两个假设条件成立. 进一步假设 $E(X) = 0$, Zhang 等 (2012b) 在累积分布降维方法的基础上给出了如下的等式:

$$\begin{aligned} \Sigma_X^{-1}E\left[XI\{\widetilde{Y} \leqslant y\}\right] &= \Sigma_X^{-1}E\left[I\{\widetilde{Y} \leqslant y\}E\left[X|(Y,U)\right]\right] \\ &= \Sigma_X^{-1}E\left[I\{\widetilde{Y} \leqslant y\}E\left(X|Y\right)\right] \\ &= B(B^{\mathrm{T}}\Sigma_X B)^{-1}E\left[I\{\widetilde{Y} \leqslant y\}B^{\mathrm{T}}X\right] \\ &\subseteq \mathcal{S}(B). \end{aligned} \tag{8.4.2}$$

记 $\widetilde{M}(y) = E[XI\{\widetilde{Y} \leqslant y\}]$, 由式 (8.4.2) 可构造如下的降维矩阵:

$$\Lambda_{\mathrm{DCU}} = \Sigma_X^{-1}E\left[\widetilde{M}(\widetilde{Y}^*)\widetilde{M}^{\mathrm{T}}(\widetilde{Y}^*)\right]\Sigma_X^{-1}, \tag{8.4.3}$$

其中 $\widetilde{Y}^*$ 是与响应变量 $\widetilde{Y}$ 相互独立, 且有相同的分布函数. 特殊地, 当结构维数 $k = 1$ 时 (即 $B = \beta_1$), 由式 (8.4.3) 定义的降维矩阵 Λ_{DCU} 可简化为

$$\Lambda_{\mathrm{DCU}} = E_{\widetilde{Y}^*}\left[\widetilde{m}^2(\widetilde{Y}^*)\right]\beta_1\beta_1^{\mathrm{T}},$$

其中

$$\widetilde{m}(\widetilde{Y}^*) = \frac{E_{(X,\widetilde{Y})}\left[\beta_1^{\mathrm{T}} X F_{\widetilde{Y}|\beta_1^{\mathrm{T}}X}(\widetilde{Y}^*|\beta_1^{\mathrm{T}}X)\right]}{\beta_1^{\mathrm{T}}\Sigma_X\beta_1},$$

且 $E_{(X,\widetilde{Y}^*)}(\cdot)$ 表示对随机向量 $(X,\widetilde{Y}^*)$ 求数学期望. 式 (8.4.3) 定义的降维矩阵 Λ_{DCU} 中 X 是不可观测的, 而 $\widetilde{Y}$ 是可以观测的. 因此, 可用 Cui 等 (2009) 提出的直接估计方法首先将协变量 X 估计出来, 然后对降维矩阵 Λ_{DCU} 进行估计. 假设 $\{(\widetilde{Y}_i,\widetilde{X}_i,U_i), i=1,\cdots,n\}$ 是来自于模型 (8.4.1) 的一组 i.i.d.的随机样本, 介绍具体估计步骤如下.

步骤 1 注意到 $E(\widetilde{X}_r|U)=\psi_r(U)E(X_r)$, $E(X_r)=E(\widetilde{X}_r)$, $r=1,\cdots,p$. 则有

$$\psi_r(u)=\frac{E(\widetilde{X}_r|U=u)}{E(\widetilde{X}_r)},\quad r=1,\cdots,p. \tag{8.4.4}$$

由式 (8.4.4), 可用非参数 Nadaraya-Watson 核光滑方法来估计 $g_{X_r}(u)=E(\widetilde{X}_r|U=u)$, $r=1,\cdots,p$, 定义如下:

$$\hat{g}_{X_r}(u)=\frac{\sum_{i=1}^{n}K_h(U_i-u)\widetilde{X}_{ir}}{\sum_{i=1}^{n}K_h(U_i-u)},\quad r=1,\cdots,p,$$

其中 $K_h(\cdot)=K(\cdot/h)/h$, $K(\cdot)$ 是核函数, h 是窗宽. 结合矩估计 $\widehat{E[\widetilde{X}_r]}=\frac{1}{n}\sum_{i=1}^{n}\widetilde{X}_{ir}$ 和 Nadaraya-Watson 核光滑估计 $\hat{g}_{X_r}(u)$, 可得 $\hat{\psi}_r(U_i)=\frac{\hat{g}_{X_r}(U_i)}{\widehat{E[\widetilde{X}_r]}}$, 并有 $\hat{X}_{ir}=\frac{\widetilde{X}_{ir}}{\hat{\psi}_r(U_i)}$, $r=1,\cdots,p$, $i=1,\cdots,n$.

步骤 2 记 $\hat{X}_i=(\hat{X}_{i1},\cdots,\hat{X}_{ip})^{\mathrm{T}}$. 基于样本 $\{(\widetilde{Y}_i,\hat{X}_i,U_i), i=1,\cdots,n\}$, 分别定义 $\widetilde{M}(y)$ 和 Σ_X 的估计如下:

$$\widehat{\widetilde{M}}(y)=\frac{1}{n}\sum_{j=1}^{n}\hat{X}_j I\{\widetilde{Y}_j\leqslant y\}$$

和

$$\hat{\Sigma}_X=\frac{1}{n}\sum_{i=1}^{n}\left(\hat{X}_i-\overline{X}\right)\left(\hat{X}_i-\overline{X}\right)^{\mathrm{T}},\qquad \overline{X}=\frac{1}{n}\sum_{i=1}^{n}\hat{X}_i.$$

步骤 3 由步骤 2 得到的估计 $\widehat{\widetilde{M}}(y)$ 和 $\hat{\Sigma}_X$ 后, 可定义降维矩阵 Λ_{DCU} 估计为

$$\hat{\Lambda}_{\mathrm{DCU}}=\hat{\Sigma}_X^{-1}\frac{1}{n}\sum_{i=1}^{n}\widehat{\widetilde{M}}\left(\widetilde{Y}_i\right)\widehat{\widetilde{M}}^{\mathrm{T}}\left(\widetilde{Y}_i\right)\hat{\Sigma}_X^{-1}.$$

得到降维矩阵 Λ_{DCU} 的估计 $\hat{\Lambda}_{\mathrm{DCU}}$ 后, 对估计的降维矩阵 $\hat{\Lambda}_{\mathrm{DCU}}$ 进行如下的特征值分解:

$$\hat{\Lambda}_{\mathrm{DCU}}\hat{\eta}_s = \hat{\lambda}_s\hat{\eta}_s, \quad \hat{\lambda}_1 \geqslant \hat{\lambda}_2 \geqslant \cdots \geqslant \hat{\lambda}_k > 0.$$

最后, 将前 k 个最大特征值 $\hat{\lambda}_1, \hat{\lambda}_2, \cdots, \hat{\lambda}_k$ 所对应的特征向量 $\hat{\eta}_1, \hat{\eta}_2, \cdots, \hat{\eta}_k$ 所张成的空间 $\mathcal{S}(\hat{\eta}_1, \hat{\eta}_2, \cdots, \hat{\eta}_k)$ 作为降维空间 $\mathcal{S}(B)$ 的估计.

为了得到估计量的渐近结果, 首先给出下面的正则条件.

(C8.4.1) 失真函数 $\psi_r(u)$, $g_{X_r}(u)$ 和混淆变量 U 的密度函数 $f_U(u)$ 都大于某个正常数, 并且有三阶连续导数, $r = 1, \cdots, p$. 进一步假定三阶导数满足下面的 Lipschiz 连续条件: 存在常数 $c > 0$, 使得对任意的 $\epsilon \in \Theta$ (其中 Θ 为包含原点的一个邻域), 都有

$$\begin{aligned} &|f_U^{(3)}(u+\epsilon) - f_U^{(3)}(u)| < c|\epsilon|, \\ &|g_{X_r}^{(3)}(u+\epsilon) - g_{X_r}^{(3)}(u)| < c|\epsilon|, \qquad r = 1, \cdots, p. \end{aligned}$$

(C8.4.2) 核函数 $K(\cdot)$ 是定义在闭区间 $[-1, 1]$ 上的连续且关于原点对称的函数, 同时要求 $K(\cdot)$ 的一阶导数有界, 并有

$$\int_{-1}^{1} K(u)\mathrm{d}u = 1, \quad \int_{-1}^{1} u^i K(u)\mathrm{d}u = 0, \quad i = 1, 2, 3, \quad \int_{-1}^{1} u^4 K(u)\mathrm{d}u \neq 0.$$

(C8.4.3) 窗宽 h 满足: 当 $n \to \infty$ 时, 有 $h \to 0$ 和 $\log n/\sqrt{nh^2} \to 0$.

(C8.4.4) 假设 $E(X_r) \neq 0$, $r = 1, \cdots, p$.

定理 8.4.1 假设正则条件(C8.4.1)~ 条件(C8.4.4)成立, 当 $n \to \infty$ 时, 则有

(1) $\hat{\Lambda}_{\mathrm{DCU}} \xrightarrow{P} \Lambda_{\mathrm{DCU}}$;

(2) 令 e 是一个长度为 p 的单位列向量, $e \in \mathcal{S}^{\perp}(B)$, 其中 $\mathcal{S}^{\perp}(B)$ 是 $\mathcal{S}(B)$ 的正交补空间. 步骤 3 得到的特征向量 $\hat{\eta}_s$ 满足: $e^{\mathrm{T}}\hat{\eta}_s$ 以 $\sqrt{n}$ 的速度收敛到均值为 0 的正态分布, $s = 1, \cdots, k$.

定理 8.4.1 的结论 (2) 可以用来检验特征向量 $\hat{\eta}_1, \cdots, \hat{\eta}_k$ 中的分量是否显著, 从而达到变量选择的目的.

8.4.3 数值模拟

本节通过模拟研究比较所提估计方法在有限样本下的完成情况. 在模拟研究中, 考虑四种样本量大小 $n = 100, 200, 300$ 和 400. 模拟数据从下面五个模型中产生:

$$Y = \frac{1}{2}X_1 + \frac{1}{2}X_2 + \frac{1}{2}X_3 + \frac{1}{2}X_4 + \varepsilon, \tag{8.4.5}$$

$$Y = \exp\left(\frac{2}{\sqrt{13}}X_1 - \frac{3}{\sqrt{13}}X_2 + \varepsilon\right), \tag{8.4.6}$$

$$Y = \sin\left(\frac{1}{\sqrt{2}}X_1 - \frac{1}{\sqrt{2}}X_2 + \frac{1}{3}\varepsilon\right), \tag{8.4.7}$$

$$Y = \log(|X_1 - 4|) + 0.5\varepsilon, \tag{8.4.8}$$

$$Y = \frac{X_1 - 5}{\frac{1}{\sqrt{2}}X_2 - \frac{1}{\sqrt{2}}X_3 + 1} + 0.5\varepsilon, \tag{8.4.9}$$

其中协变量 X_j 来自于 $t(5)+6$ 的分布, $j=1,\cdots,p$. 模型误差 ε 从正态分布 $N(0,1.44)$ 中产生, 且与 X_j 独立, $j=1,\cdots,p$. 混淆变量 U 来自于区间 $(1,6)$ 上的均匀分布. 响应变量 Y 的失真函数取为 $\phi(U)=\dfrac{(U+1)^2}{22.3333}$, 协变量 X 的失真函数取为 $\psi_r(U)=\dfrac{U+r}{3.5+r}$, $r=1,\cdots,p$. 核函数取高阶核函数 $K(t)=\dfrac{15}{32}(3-7t^2)I(|t|\leqslant 1)$, 采用交错核实方法选取窗宽 h, 然后对 $\{\psi_i(U), i=1,\cdots,p\}$ 进行估计.

模拟结果用式 (8.3.20) 定义的迹相关准则 (Férre, 1998) 对步骤 3 中得到的特征向量张成空间 $\mathcal{S}(\hat{\eta}_1,\cdots,\hat{\eta}_k)$ 和真实空间 $\mathcal{S}(B)$ 的距离进行评价. 为了比较, 考虑协变量的维数为 $p=10,20,30$ 三种情形下迹相关系数 $R^2(k)$ 的表现. 本例重复模拟试验进行了 300 次, 模拟结果见表 8.4.1.

从表 8.4.1 中的模拟结果可以看出: 迹相关系数 $R^2(k)$ 的表现和协变量 X 的维数和样本量 n 密切相关. 样本量 n 固定时, 如果协变量 X 维数比较大 (如 $p=30$), $R^2(k)$ 的表现较差. 但当维数 p 固定, 样本量 n 不断增加时, $R^2(k)$ 改善效果很明显. 当 $p=30$, 在样本量 n 达到 400 的时候, $R^2(k)$ 至少为 0.74. 说明本节的方法在大样本情形下, 估计结果的表现会越来越好.

表 8.4.1 迹相关系数 $R^2(k)$ 的平均均值和标准差

p	模型	结果	$n=100$	$n=200$	$n=300$	$n=400$
$p=10$	模型 (8.4.5)	均值	0.8299	0.9216	0.9498	0.9719
		标准差	0.0867	0.0403	0.0242	0.0137
	模型 (8.4.6)	均值	0.8738	0.9300	0.9487	0.9629
		标准差	0.0619	0.0377	0.0260	0.0167
	模型 (8.4.7)	均值	0.8368	0.9171	0.9409	0.9567
		标准差	0.1161	0.0591	0.0405	0.0280
	模型 (8.4.8)	均值	0.8173	0.8943	0.9341	0.9484
		标准差	0.1192	0.0636	0.0362	0.0307
	模型 (8.4.9)	均值	0.7175	0.8226	0.8833	0.9091
		标准差	0.1070	0.0711	0.0434	0.0331

续表

p	模型	结果	$n=100$	$n=200$	$n=300$	$n=400$
$p=20$	模型 (8.4.5)	均值	0.7185	0.8375	0.8954	0.9236
		标准差	0.1008	0.0538	0.0376	0.0261
	模型 (8.4.6)	均值	0.7423	0.8724	0.9088	0.9300
		标准差	0.0864	0.0443	0.0317	0.0216
	模型 (8.4.7)	均值	0.6998	0.8396	0.8864	0.9177
		标准差	0.1431	0.0661	0.0578	0.0375
	模型 (8.4.8)	均值	0.6436	0.8018	0.8566	0.8956
		标准差	0.1577	0.0916	0.0716	0.0509
	模型 (8.4.9)	均值	0.5071	0.6843	0.7657	0.8215
		标准差	0.1153	0.0764	0.0706	0.0495
$p=30$	模型 (8.4.5)	均值	0.5222	0.7609	0.8397	0.8844
		标准差	0.1240	0.0750	0.0418	0.0307
	模型 (8.4.6)	均值	0.6034	0.8084	0.8671	0.9012
		标准差	0.1171	0.0536	0.0403	0.0264
	模型 (8.4.7)	均值	0.5639	0.7517	0.8367	0.8849
		标准差	0.1683	0.1054	0.0720	0.0375
	模型 (8.4.8)	均值	0.5107	0.6995	0.7989	0.8515
		标准差	0.1711	0.1185	0.0698	0.0584
	模型 (8.4.9)	均值	0.3728	0.5702	0.6800	0.7429
		标准差	0.1067	0.0852	0.0658	0.0553

参考文献

柴根象, 洪圣岩. 1995. 半参数回归模型. 合肥: 安徽教育出版社.

柴根象, 孙平. 1995. 半参数回归模型的二阶段估计的渐近正态性. 应用数学学报, 18(3): 354–363.

柴根象, 徐克军. 1997. 半参数回归模型的小波光滑. 第五届中国统计讨论会议论文集.

陈明华. 1998a. 固定设计下半参数回归模型估计的相合性. 高校应用数学学报, 13(3): 301–310.

陈明华. 1998b. 固定设计下半参数回归模型估计的收敛速度. 应用概率统计, 14(2): 149–158.

崔恒建. 1995a. EV 模型中参数 M 估计的渐近正态性. 科学通报, 40(22): 2031–2033.

崔恒建. 1995b. 半参数 EV 模型的参数估计理论. 科学通报, 40(16): 1444–1447.

崔恒建. 1997. EV 模型中广义最小一乘估计的渐近性质. 中国科学 (A 辑), 27(2): 119–131.

崔恒建. 2004. 有重复观测的部分 EV 模型的参数估计. 中国科学 (A 辑), 34(4): 467–482.

崔恒建, 李勇, 秦怀振. 1998. 非线性半参数 EV 回归模型的估计理论. 科学通报, 23: 2493–2497.

冯三营, 李高荣, 薛留根, 等. 2010a. 非线性半参数 EV 模型的经验似然置信域. 高校应用数学学报, 25(1): 53–63.

冯三营, 牛惠芳, 薛留根. 2010b. 变系数 EV 模型的局部纠偏经验似然. 应用概率统计, 26(4): 427–436.

冯三营, 裴丽芳, 薛留根. 2011. 非参数部分带有测量误差的部分线性变系数模型的经验似然推断. 系统科学与数学, 31(12): 1652–1663.

冯三营, 薛留根. 2012. 非线性半参数 EV 模型的最大经验似然估计. 数学物理学报, 32(4): 729–743.

冯三营, 薛留根, 李高荣. 2007. 非线性 EV 回归模型中参数的经验似然置信域. 应用数学, 20(2): 400–407.

高集体, 陈希孺, 赵林诚. 1994. 部分线性模型中估计的渐近正态性. 数学学报, 37(2): 256–268.

高集体, 洪圣岩, 梁华. 1995. 部分线性模型中估计的收敛速度. 数学学报, 38(5): 658–669.

胡舒合. 1994. 固定设计下半参数回归模型估计的强相合性. 数学学报, 37(3): 393–401.

李高荣, 冯三营, 薛留根. 2007. 核实数据下删失线性 EV 模型的经验似然推断. 数理统计与管理, 26(5): 782–791.

李高荣, 薛留根. 2007a. 经验 Cressie-Read 统计量与广义矩方法估计. 工程数学学报, 28(6): 1031–1041.

李高荣, 薛留根. 2007b. 弱相依数据中的分组经验 Cressie-Read 似然方法. 应用概率统计, 23(2): 197–206.

李高荣, 薛留根, 冯三营. 2006. 部分线性 EV 回归模型中的极大经验似然估计. 北京工业大学学报, 32(7): 667–672.

李高荣, 薛留根, 冯三营. 2009. 核实数据下响应变量缺失的线性 EV 模型的经验似然推断.

系统科学与数学, 29(1): 94–108.
李高荣, 杨宜平. 2015. 纵向数据半参数模型. 北京: 科学出版社.
刘强. 2010. 缺失数据下非线性半参数 EV 模型的估计. 系统科学与数学, 9: 1236–1250.
钱伟民, 柴根象. 1999. 半参数回归模型小波估计的强逼近. 中国科学, 29(3): 233–240.
钱伟民, 柴根象, 蒋凤英. 2000. 半参数回归模型的误差方差的小波估计. 数学年刊, 21A(3): 341–350.
田萍. 2006. 纵向数据半参数回归模型的估计理论. 北京: 北京工业大学博士学位论文.
王启华. 2004. 经验似然统计推断方法发展综述. 数学进展, 33(2): 141–151.
王启华. 1995. 随机截断下半参数回归模型的相合估计. 中国科学, 25(8): 819–832.
王启华, 史宁中, 耿直. 2010. 现代统计研究基础. 北京: 科学出版社.
王启华, 郑忠国. 1997. 随机删失下半参数回归模型中估计的渐近性质. 中国科学, 27(7): 583–594.
魏传华, 吴喜之. 2008. 部分线性变系数模型的 Profile Lagrange 乘子检验. 系统科学与数学, 28(4): 416–424.
薛留根. 1999. 随机删失下半参数回归模型的估计理论. 数学年刊, 20(6): 745–754.
薛留根. 2002. 半参数回归模型随机加权 M 估计的强逼近. 应用数学学报, 25(4): 591–603.
薛留根. 2003. 半参数回归模型中小波估计的随机加权逼近速度. 应用数学学报, 26(1): 11–25.
薛留根. 2004. 半参数回归模型中误差方差估计之分布的非一致收敛速度. 应用数学学报, 27(2): 379–381.
薛留根. 2005a. 半参数回归模型中误差方差估计的 Berry-Esseen 界. 数学学报, 48(1): 157–170.
薛留根. 2005b. 非线性 EV 回归模型中参数估计的渐近性质. 数学年刊, 26: 351–360.
薛留根. 2006. 核实数据下非线性半参数 EV 模型的经验似然推断. 数学学报, 49: 145–154.
薛留根. 2012a. 现代统计模型. 北京: 科学出版社.
薛留根. 2012b. 单指标模型的统计推断. 数理统计与管理, 31(1): 55–78.
薛留根. 2012c. 单指标模型的统计推断. 数理统计与管理, 31(2): 226–246.
薛留根. 2015. 现代非参数统计. 北京: 科学出版社.
印赞华. 2013. 非参数测量误差模型的统计推断. 长春: 东北师范大学.
张日权, 卢一强. 2004. 变系数模型. 北京: 科学出版社.
赵培信, 薛留根. 2011. 协变量调整回归模型的经验似然推断. 数学物理学报, 31A(2): 360–368.
Ahmad I, Leelahanon S, Li Q. 2005. Efficient estimation of a semiparametric partially linear varying coefficient model. The Annals of Statistics, 33(1): 258–283.
Akaike H. 1973. Maximum likelihood identification of Gaussian autoregressive moving average models. Biometrika, 22: 203–217.
Antoniadis A. 1997. Wavelets in statistics: a review (with discussion). Journal of the Italian Statistical Society, 6: 97–144.
Apanasovich T V, Carroll R J. 2009. SIMEX and standard error estimation in semipara-

metric measurement error models. Electronic Journal of Statistics, 3: 318–348.

Azzalini A, Bowman A. 1993. On the use of nonparametric regression for checking linear relationships. Journal of the Royal Statistical Society, Series B, 55: 549–557.

Bai Y, Hu J, You J. 2015. Panel data partially linear varying coefficient model with errors correlated in space and time. Statistica Sinica, 25, 275–294.

Bates D, Watts D. 1988. Nonlinear Regression Analysis and Its Applications. New York: John Wiley & Sons.

Breiman L. 1996. Heuristics of instability and stabilization in model selection. The Annals of Statistics, 24: 2350–2383.

Bura E, Cook R D. 2001a. Extending sliced inverse regression: the weighted chi-squared test. Journal of the American Statistical Association, 96: 996–1003.

Bura E, Cook R D. 2001b. Estimating the structural dimension of regressions via parametric inverse regression. Journal of the Royal Statistical Society, Series B, 63: 393–410.

Buzas J S. 1997. Instrumental variable estimation in nonlinear measurement error models. Communications in Statistics—Theory and Methods, 26(12): 2861–2877.

Cai Z W. 2002. Two-step likelihood estimation procedure for varying-coefficient models. Journal of Multivariate Analysis, 82: 189–209.

Cai Z W. 2003. Local quasi-likelihood approach to varying-coefficient discrete-valued time series models. Journal of Nonparametric Statistics, 15(6): 693–711.

Cai Z W, Fan J Q, Li R Z. 2000. Efficient estimation and inferences for varying-coefficient models. Journal of the American Statistical Association, 95: 888–902.

Carroll R J, Fan J Q, Gijbels I, et al. 1997. Generalized partially linear single-index models. Journal of the American Statistical Association, 92: 477–489.

Carroll R J, Li K C. 1992. Measurement error regression with unknown link: dimension reduction and data visualization. Journal of the American Statistical Association, 87: 1040–1050.

Carroll R J, Lombard F, Küchenhoff H, et al. 1996. Asymptotics for the SIMEX estimator in structural measurement error models. Journal of the American Statistical Association, 91: 242–250.

Carroll R J, Maca J, Ruppert D. 1999. Nonparametric regression in the presence of measurement error. Biometrika, 86: 541–554.

Carroll R J, Ruppert D. 1988. Transformation and Weighting in Regression. New York: Chapmen & Hall/CRC.

Carroll R J, Ruppert D, Stefanski L A. 1995. Measurement Error in Nonlinear Models. London: Chapmen & Hall/CRC.

Carroll R J, Ruppert D, Stefanski L A, Crainiceanu C M. 2006. Measurement Error in Nonlinear Models: A Modern Perspective. London: Chapmen & Hall/CRC.

Chan L K, Mak T K. 1985. On the polynomial functional relationship. Journal of the

Royal Statistical Society: Series B, 47: 510–518.

Chang Z Q, Xue L G, Zhu L X. 2010. On an asymptotically more efficient estimation of the single-index model. Journal of Multivariate Analysis, 101: 1898–1901.

Chen H. 1988. Convergence rates for parametric components in a partly linear model. The Annals of Statistics, 16: 136–146.

Chen S X, Cui H J. 2003. An extended empirical likelihood for generalized linear models. Statistica Sinica, 13: 69–81.

Chen T, Parker T. 2014. Semiparametric efficiency for partially linear single-index regression models. Journal of Multivariate Analysis, 130: 376–386.

Chen X, Cui H J. 2009. Empirical likelihood for partially linear single-index errors-in-variables model. Communications in Statistics–Theory and Methods, 38: 2498–2514.

Chen X, Cui H J. 2012. Empirical likelihood inference for parameters in a partially linear errors-in-variables model. Statistics, 46(6): 745–757.

Chen X, Zou C L, Cook D R. 2010. Coordinate-independent sparse sufficient dimension reduction and variable selection. The Annals of Statistics, 38: 3696–3723.

Chiang C T, Rice J A, Wu C O. 2001. Smoothing spline estimation for varying coefficient models with repeatedly measured dependent variables. Journal of the American Statistical Association, 96(454): 605–619.

Chiou J M, Müller H G. 1999. Nonparametric quasi-likelihood. The Annals of Statistics, 27(1): 36–64.

Cleveland W S, Grosse E, Shyu W M. 1992. Local regression models. Statistical models in S, 309–376.

Cook J, Stefanski L A. 1994. Simulation-extrapolation method in parametric measurement error models. Journal of the American Statistical Association, 89: 1314–1328.

Cook R D. 1998a. Regression Graphics: Ideas for Studying Regressions Through Graphics. New York: John Wiley & Sons.

Cook R D. 1998b. Principal Hessian directions revisited. Journal of the American Statistical Association, 93: 84–94.

Cook R D. 2000. SAVE: A method for dimension reduction and graphics in regression. Communications in Statistics - Theory and Methods, 29: 2109–2121.

Cook R D, Li B. 2004. Determining the dimension of iterative Hessian transformation. The Annals of Statistics, 32: 2501–2531.

Craven P, Wahba G. 1979. Smoothing noisy data with spline functions: estimating the correct degree of smoothing by the method of generalized cross-validation. Numerische Mathematik, 31: 337–403.

Cui H J, Chen S X. 2003. Empirical likelihood confidence region for parameter in the errors-in-variables models. Journal of Multivariate Analysis, 84(1): 101–115.

Cui H J, Kong E F. 2006. Empirical likelihood confidence regions for semi-parametric

errors-in-variables models. Scandinavian Journal of Statistics, 33: 153–168.

Cui H J, He X M, Zhu L X. 2002. On regression estimators with de-noised variables. Statistica Sinica, 12(4): 1191–1205.

Cui H J, Hu T. 2011. On nonlinear regression estimator with denoised variables. Computational Statistics and Data Analysis, 55(2): 1137–1149.

Cui X. 2008. Statistical Analysis of Two Types of Complex Data and Its Associated Model. Jinan: Ph.D. Thesis, Shandong University.

Cui X, Guo W S, Lin L, Zhu L X. 2009. Covariate-adjusted nonlinear regression. The Annals of Statistics, 37: 1839–1870.

Cui X, Härdle W, Zhu L X. 2011. Generalized single-index models: the EFM approach. The Annals of Statistics, 39: 1658–1688.

Delaigle A, Fan J Q, Carroll R J. 2009. A design-adaptive local polynomial estimation for the errors-in-variables problem. Journal of the American Statistical Association, 104: 348–359.

Delaigle A, Hall P. 2008. Using SIMEX for smoothing parameter choice in errors-in-variables problems. Journal of the American Statistical Association, 130: 280–287.

Delaigle A, Hall P, Meister A. 2008. On deconvolution with repeated measurements. The Annals of Statistics, 36: 665–685.

Delaigle A, Meister A. 2007. Nonparametric regression estimation in the heteroscedastic errors-in-variables problem. Journal of the American Statistical Association, 102(480): 1416–1426.

Dette H. 1999. A consistent test for the functional form of a regression based on a difference of variance estimators. The Annals of Statistics, 27(3): 1012–1040.

Ding X B, Zhou X H, Wang Q H. 2015. A partially linear single-index transformation model and its nonparametric estimation. Canadian Journal of Statistics, 43: 97–117.

Eaton M L. 1986. A characterization of spherical distributions. Journal of Multivariate Analysis, 34: 439–446.

Eaton M, Tyler D. 1991. On wielandt's inequality and its applications to the asymptotic distribution of the eigenvalues of a random symmetric matrix. The Annals of Statistics, 19: 260–271.

Efron B. 1979. Bootstrap methods: another look at the Jackknife. The Annals of Statistics, 7(1): 1–26.

Efron B, Hastie T, Johnstone L, Tibshirani R. 2004. Least angle regression. The Annals of Statistics, 32: 407–499.

Efron B, Tibshirani R. 1993. An Introduction to the Bootstrap. London: Chapmen & Hall/CRC.

Engle R F, Granger C W J, Rice J, et al. 1986. Semiparametric estimates of the relation between weather and electricity scales. Journal of the American Statistical Association,

81: 310–320.

Eubank R L, Huang C F, Buchanan R J. 2004. Smoothing spline estimation in varying-coefficient models. Journal of the Royal Statistical Society: Series B, 66: 653–667.

Eubank R L, Kambour E L, Kim J T, et al. 1998. Estimation in partially linear models. Computational Statistics & Data Analysis, 29: 27–34.

Fan J Q. 1993. Local linear regression smoothers and their minimax efficiencies. The Annals of Statistics, 21(1): 196–216.

Fan J Q. 1996. Test of significance based on Wavelet thresholding and Neyman's truncation. Journal of the American Statistical Association, 91: 674–688.

Fan J Q. 1997. Comment on "Wavelets in statistics: a review" by Antoniadis A. Journal of the Italian Statistical Society, 6: 131–138.

Fan J Q, Gijbels I. 1996. Local Polynomial Modeling and Its Application. London: Chapman & Hall.

Fan J Q, Huang L S. 2001. Goodness-of-fit tests for parametric regression models. Journal of the American Statistical Association, 96: 640–652.

Fan J Q, Huang T. 2005. Profile likelihood inferences on semiparametric varying-coefficient partially linear models. Bernoulli, 11(6): 1031–1057.

Fan J Q, Li R Z. 2001. Variable selection via nonconcave penalized likelihood and its oracle properties. Journal of the American Statistical Association, 96: 1348–1360.

Fan J Q, Lv J C. 2008. Sure independence screening for ultra-high dimensional feature space (with discussion). Journal of the Royal Statistical Society, Series B, 70: 849–911.

Fan J Q, Peng H. 2004. Nonconcave penalized likelihood with a diverging number of parameters. The Annals of Statistics, 32: 928–961.

Fan J Q, Truong Y K. 1993. Nonparametric regression with errors in variables. The Annals of Statistics, 21(4): 1900–1925.

Fan J Q, Yao Q W, Tong H. 1996. Estimation of conditional densities and sensitivity measures in nonlinear dynamical systems. Biometrika, 83: 189–196.

Fan J Q, Zhang C M, Zhang J. 2001. Generalized likelihood ratio statistics and Wilks phenomenon. The Annals of Statistics, 29: 153–193.

Fan J Q, Zhang J T. 2000. Two-step estimation of functional linear models with applications to longitudinal data. Journal of the Royal Statistical Society, Series B, 62: 303–322.

Fan J Q, Zhang W Y. 1999. Statistical estimation in varying coefficient models. The Annals of Statistics, 27: 1491–1518.

Fan J Q, Zhang W Y. 2000. Simultanenous confidence bands and hypothesis testing in varying-coefficient models. Scandinavian Journal of Statistics, 27: 715–731.

Fan J Q, Zhang W Y. 2008. Statistical methods with varying coefficient models. Statistics and Its Inference, 1: 179–195.

Fang K T, Kotz S, Ng K W. 1990. Symmetric Multivariate and Related Distributions.

London: Chapmen & Hall/CRC.

Feng L, Zou C, Wang Z, et al. 2015. Robust spline-based variable selection in varying coefficient model. Metrika, 78(1): 85–118.

Feng S Y, Li G R, Zhang J H. 2014. Efficient statistical inference for partially nonlinear errors-in-variables models. Acta Mathematica Sinica, English Series, 30(9): 1606–1620.

Feng S Y, Xue L G. 2014. Bias-corrected statistical inference for partially linear varying coefficient errors-in-variables models with restricted condition. Annals of the Institute of Statistical Mathematics, 66: 121–140.

Férre L. 1998. Determining the dimension in sliced inverse regression and related methods. Journal of the American Statistical Association, 93: 132–140.

Firth D. 1987. On the efficiency of quasi-likelihood. Biometrika, 74: 233–245.

Foster D P, George E I. 1994. The risk inflation criterion for multiple regression. The Annals of Statistics, 22: 1947–1975.

Frank I E, Friedman J H. 1993. A statistical view of some chemometrics regression tools (with discussion). Technometrics, 35: 109–148.

Friedman J H, Stuetzle W. 2001. Projection pursuit regression. Journal of the American Statistical Association, 76: 817–823.

Fuller W A. 1987. Measurement Error Models. New York: John Wiley & Sons.

Fung W K, He X M, Liu L, Shi P D. 2002. Dimension reduction based on canonical correlation. Statistica Sinica, 12: 1093–1113.

Gallant A R. 1987. Nonlinear Regression Models. New York: John Wiley & Sons.

Gao J T, Zhao L C. 1993. Adaptive estimation in partly linear regression models. Science in China Series A: Mathematics, 23(1): 14–27.

Gasser T, Müller H G. 1979. Kernel estimation of regression function // Smoothing Techniques for Curve Estimation, Lecture Notes in Mathematics, 757: 23–68. New York: Springer-Verlag.

Green P, Jennison C, Seheult A. 1985. Analysis of Field Experiments by Least Squares Smoothing. Journal of the Royal Statistical Society, Series B, 47: 299–315.

Green P, Silverman B W. 1994. Nonparametric Regression and Generalized Linear Models: a Roughness Penalty Approach. New York: Chapmen & Hall/CRC.

Gu J, Liang, Z. 2014. Testing cointegration relationship in a semiparametric varying coefficient model. Journal of Econometrics, 178: 57–70.

Guo X, Niu C Z, Yang Y P, et al. 2015. Empirical likelihood for single index model with missing covariates at random. Statistics: A Journal of Theoretical and Applied Statistics, 49: 588–601.

Hall P. 1992. The Bootstrap and Edgeworth Expansion. New York: Springer-Verlag.

Hall P, Li K C. 1993. On almost linearity of low dimensional projection from high dimensional data. The Annals of Statistics, 21: 867–889.

Hamilton S A, Truong Y K. 1997. Local linear estimation in partly models. Journal of Multivariate Analysis, 60: 1–19.

Hammer S M, Katzenstein D A, Hughes M D, et al. 1996. A trial comparing nucleoside monotherapy with combination therapy in hiv-infected adults with cd4 cell counts from 200 to 500 per cubic millimeter. New England Journal of Medicine, 335: 1081–1090.

Hansen L P. 1982. Large sample properties of generalized method of moments estimators. Econometrica, 50: 1029–1054.

Härdle W. 1990. Applied Nonparametric Regression. Boston: Cambridge University Press.

Härdle W, Liang H, Gao J T. 2000. Partially Linear Models. Heidelberg: Physica Verlag.

Härdle W, Mammen E. 1993. Comparing nonparametric versus parametric regression fits. The Annals of Statistics, 21(4), 1926–1947.

Härdle W, Mammen E, Müller M. 1998. Testing parametric versus semiparametric modeling in generalized linear models. Journal of American Statistical Association, 93: 1461–1474.

Härdle W, Hall P, Ichimura H. 1993. Optimal smoothing in single-index models. The Annals of Statistics, 21: 157–178.

Härdle W, Stoker T M. 1989. Investigating smooth multiple regression by the method of average derivatives. Journal of the American Statistical Association, 84: 986–995.

Harrison D, Rubinfeld D L. 1978. Hedonic prices and the demand for clean air. Journal of Environmental Economics and Management, 5: 81–102.

Hart J D. 1997. Nonparametric Smoothing and Lack-of-Fit Tests. Springer Series in Statistics, New York: Springer-Verlag.

Hastie T, Tibshirani R. 1990. Generalized Additive Models. London: Chapmen & Hall/CRC.

Hastie T, Tibshirani R. 1993. Varying-coefficient models. Journal of the Royal Statistical Society, Series B (Methodological), 55(4): 757–796.

He X, Liang H. 2000. Quantile regression estimates for a class of linear and partially linear errors-in-variables models. Statistica Sinica, 10: 129–140.

He X, Zhu Z Y, Fung W K. 2002. Estimation in a semiparametric model for longitudinal data with unspecified dependence structure. Biometrika, 89: 579–590.

Heaney R P. 2003. Normalizing calcium intake: projected population effects for body weight. Journal of Nutrition, 133: 268S–270S.

Heaney R P, Recker R R, Stegman M R, et al. 1989. Calcium absorption in women: relationships to calcium intake, estrogen status, age. Journal of Bone and Mineral Research, 4: 469–475.

Heckman N E. 1986. Spline smoothing in a partly linear model. Journal of the Royal Statistical Society, Series B, 48: 244–248.

Homg Z P, Lian H. 2012. Time-varying coefficient estimation in differential equation models with noisy time-varying covariates. Journal of Multivariate Analysis, 103(1): 58–67.

Honda T. 2004. Quantile regression in varying coefficient models. Journal of Statistical Planning and Inference, 121: 113–125.

Hooper J. 1959. Simultaneous equations and canonical correlation theory. Econometrica, 27: 245–256.

Hoover D R, Rice J A, Wu C O, et al. 1998. Nonparametric smoothing estimates of time-varying coefficient models with longitudinal data. Biometrika, 85: 809–822.

Horowitz J L. 2009. Semiparametric and Nonparametric Methods in Econometrics. New York: Springer.

Horowitz J L, Härdle W. 2009. Direct semiparametric estimation of single-index models with discrete covariates. Journal of American Statistical Association, 91: 1632–1640.

Hotelling H. 1936. Relations between two sets of variates. Biomatrika, 28: 321–377.

Hsiao C. 1989. Consistent estimation for some nonlinear errors-in-variables models. Journal of Econometrics, 41: 159–185.

Hsiao C, Wang L, Wang Q. 1997. Estimation of nonlinear errors-in-variables models: an approximate solution. Statistical Papers, 38(1): 1–25.

Hsing T. 1999. Nearest neighbor inverse regression. The Annals of Statistics, 27: 691–731.

Hsing T, Carroll R J. 1992. An asymptootic theory for sliced inverse regression. The Annals of Statistics, 20: 1040–1061.

Hu T, Cui H J. 2010. Robust estimates in generalised varying-coefficient partially linear models. Journal of Nonparametric Statistics, 22(6): 737–754.

Hu T, Xia Y. 2010. Adaptive semi-varying coefficient model selection. Statistica Sinica, 22: 575–599.

Hu X M, Wang Z Z, Liu F. 2009. Testing serial correlation in semiparametric varying-coefficient partially linear models. Communications in Statistics—Theory and Methods, 38(13): 2145–2163.

Huang J Z. 2003. Local asymptotics for polynomial spline regression. The Annals of Statistics, 31: 1600–1635.

Huang J Z, Wu C O, Zhou L. 2002. Varying-coefficient models and basis function approximations for the analysis of repeated measurements. Biometrika, 89(1): 111–128.

Huang J Z, Wu C O, Zhou L. 2004. Polynomial spline estimation and inference for varying coefficient models with longitudinal data. Statistica Sinica, 14(3): 763–788.

Huang Y, Wang C Y. 2001. Consistent functional methods for logistic regression with errors in covariates. Journal of the American Statistical Association, 96: 1469–1482.

Huang Z S. 2011. Empirical likelihood for generalized partially linear varyingcoefficient models. Journal of Applied Statistics, 38(6): 1265–1275.

Huang Z S. 2012. Empirical likelihood for the parametric part in partially linear errors-in-function models. Statistics & Probability Letters, 82(1): 63–66.

Huang Z S, Zhang R Q. 2009. Empirical likelihood for nonparametric parts in semipara-

metric varying-coefficient partially linear models. Statistics & Probability Letters, 79: 1798–1808.

Huber P J. 1973. Robust regression: asymptotics, conjectures and Monte Carlo. The Annals of Statistics, 1: 799–821.

Huber P J. 1981. Robust Statistics. New York: John Wiley & Sons.

Huet S. 2004. Statistical Tools for Nonlinear Regression: A Practical Guide with S-PLUS and R Examples. New York: Springer-Verlag.

Hunter D, Li R Z. 2005. Variable selection using MM algorithms. The Annals of Statistics, 33: 1617–1642.

Ichimura H. 1993. Semiparametric least squares (SLS) and weighted SLS estimation of single-index models. Journal of Econometrics, 58: 71–120.

Ip W C, Wong H, Zhang R Q. 2007. Generalized likelihood ratio test for varying-coefficient models with different smoothing variables. Computational Statistics & Data Analysis, 51: 4543–4561.

Jiang Q, Wang H, Xia Y, et al. 2013. On a principal varying coefficient model. Journal of the American Statistical Association, 108(501): 228–236.

Jin Z, Lin D Y, Wei L J, et al. 2003. Rank-based inference for the accelerated failure time model. Biometrika, 90: 341–353.

Kai B, Li R, Zou H. 2011. New efficient estimation and variable selection methods for semiparametric varying-coefficient partially linear models. The Annals of Statistics, 39(1): 305–332.

Kato T. 1983. Perturbation Theory for Linear Operation (2nd ed.). New York: Springer-Verlag.

Kauermann G, Tutz G, Bruderl J. 2005. The survival of newly founded firms: a casestudy into varying-coefficient models. Journal of the Royal Statistical Society, Series A-Statistics in Society, 168: 145–158.

Kaysen G A, Dubin J A, Müller H G, et al. 2003. Relationship among inflammation nutrition and physiologic mechanisms establishing albumin levels in hemodialysis patients. Kidney International, 61: 2240–2249.

Kim M. 2007. Quantile regression with varying coefficients. The Annals of Statistics, 35: 92–108.

Klopp O, Pensky M. 2013. Non-asymptotic approach to varying coefficient model. Electronic Journal of Statistics, 7: 454–479.

Knight K, Fu W. 2000. Asymptotics for lasso-type estimators. The Annals of Statistics, 28: 1356–1378.

Koenker R. 2005. Quantile Regression. New York: Cambridge University Press.

Kolaczyk E D. 1994. Empirical likelihood for generalized linear models. Statistica Sinica, 4: 199–218.

Lai P, Li G R, Lian H. 2013a. Quadratic inference functions for partially linear single-index models with longitudinal data. Journal of Multivariate Analysis, 118: 115–127.

Lai P, Li G R, Lian H. 2013b. Semiparametric estimation of fixed effects panel data single-index model. Statistics & Probability Letters, 83: 1595–1602.

Lam C, Fan J Q. 2008. Profile-kernel likelihood inference with diverging number of parameters. The Annals of Statistics, 36(5): 2232–2260.

Lee S, Kwon K. 2004. Pharmacokinetic-pharmacodynamic modeling for the relationship between glucose-lowering effect and plasma concentration of metformin in volunteers. Archives of Pharmacal Research, 27: 806–810.

Lee L F, Sepanski J H. 1995. Estimation of linear and nonlinear errors-in-variables models using validation data. Journal of the American Statistical Association, 90(429): 130–140.

Leng C L. 2009. A simple approach for varying-coefficient model selection. Journal of Statistical Planning and Inference, 139: 2138–2146.

Leng C L, Zhang W, Pan J. 2010. Semiparametric mean-covariance regression analysis for longitudinal data. Journal of the American Statistical Association, 105: 181–193.

Li B, Wang S. 2007. On directional regression for dimension reduction. Journal of the American Statistical Association, 102: 997–1008.

Li B, Yin X R. 2007. On surrogate dimension reduction for measurement error regression: an invariance law. The Annals of Statistics, 35: 2143–2172.

Li B, Zha H, Chiaromonte C. 2005. Contour regression: a general approach to dimension reduction. The Annals of Statistics, 33: 1580–1616.

Li D G, Chen J, Lin Z Y. 2009. Variable selection in partially time-varying coefficient models. Journal of Nonparametric Statistics, 21(5): 553–566.

Li F, Lin L, Cui X. 2010a. Covariate-adjusted partially linear regression models. Communications in Statistics-Theory and Methods, 39: 1054–1074.

Li G R, Feng S Y, Peng H. 2011a. A profile-type smoothed score function for a varying coefficient partially linear model. Journal of Multivariate Analysis, 102(2): 372–385.

Li G R, Lai P, Lian H. 2015. Variable selection and estimation for partially linear single-index models with longitudinal data. Statistics and Computing, 25: 579–593.

Li G R, Lin L, Zhu L X. 2012a. Empirical likelihood for varying coefficient partially linear model with diverging number of parameters. Journal of Multivariate Analysis, 105: 85–111.

Li G R, Peng H, Dong K, et al. 2014a. Simultaneous confidence bands and hypothesis testing in single-index models. Statistica Sinica, 24: 937–955.

Li G R, Peng H, Zhang J, et al. 2012b. Robust rank correlation based screening. The Annals of Statistics, 40: 1285–1933.

Li G R, Peng H, Zhu L X. 2011b. Nonconcave penalized M-estimation with diverging

number of parameters. Statistica Sinica, 21: 391–419.

Li G R, Xue L G. 2008. Empirical likelihood confidence region of the parameter in the partially linear errors-in-variables model. Communications in Statistics-Theory and Methods, 37: 1552–1564.

Li G R, Zhu L X, Xue L G, et al. 2010b. Empirical likelihood inference in partially linear single-index models for longitudinal data. Journal of Multivariate Analysis, 101: 718–732.

Li K C. 1991. Sliced inverse regression for dimension reduction (with discussion). Journal of American Statistical Association, 86: 316–342.

Li K C. 1992. On principal Hessian directions for data visualization and dimension reduction: another application of Stein's lemma. Journal of American Statistical Association, 87: 1025–1039.

Li K C, Aragon Y, Shedden K, et al. 2003. Dimension reduction for multivariate response data. Journal of the American Statistical Association, 98: 99–109.

Li K C, Wang J L, Chen C H. 1999. Dimension reduction for censored regression data. The Annals of Statistics, 27: 1–23.

Li L. 2007. Sparse sufficient dimension reduction. Biometrika, 94: 603–613.

Li L, Yin X. 2008. Sliced inverse regression with regularization. Biometrics, 64: 124–131.

Li N, Xu X Z, Liu X H. 2011c. Testing the constancy in varying-coefficient regression models. Metrika, 74(3): 409–438.

Li R, Liang H. 2008. Variable selection in semiparametric regression modeling. The Annals of Statistics, 36: 261–286.

Li R, Nie L. 2008. Efficient statistical inference procedures for partially nonlinear models and their applications. Biometrics, 64: 904–911.

Li T. 2000. Estimation of nonlinear errors-in-variables models: a simulated minimum distance estimator. Statistics & Probability Letters, 47(3): 243–248.

Li T. 2002. Robust and consistent estimation of nonlinear errors-in-variables models. Journal of Econometrics, 110(1): 1–26.

Li X J, Du J, Li G R, et al. 2014b. Variable selection for covariate adjusted regression model. Journal of Systems Science and Complexity, 27(6): 1227–1246.

Li Y, Lin X F, Zhang S M. 1996. The large sample theory of nonlinear errors-in-variables regression models. Progress in Natural Science, 6(6): 600–666.

Li Y, Ruppert D. 2008. On the asymptotics of penalized splines. Biometrika, 95: 415–436.

Li Y X, Zhu L X. 2007. Asymptotics for sliced average variance estimation. The Annals of Statistics, 35: 41–69.

Lian H. 2012. Variable selection for high-dimensional generalized varying-coefficient models. Statistica Sinica, 22(4): 1563–1588.

Lian H, Liang H, Carroll R M. 2015. Variance function partially linear single-index models.

Journal of the Royal Statistical Society, Series B, 77: 171–194.

Liang H. 1992. Asymptotic efficiency in semiparametric models and related topics. Beijing: Ph.D Thesis, Institute of Systems Science, Chinese Academy of Sciences.

Liang H. 2000. Asmptotic normality of parametric part in partially linear models with measurement error in the nonparametric part. Journal of Statistical Planning and Inference, 86: 51–62.

Liang H. 2009. Generalized partially linear mixed-effects models incorporating mismeasured covariates. Annals of the Institute of Statistical Mathematics, 61: 27–46.

Liang H, Härdle W, Carroll R J. 1999. Estimation in a semiparametric partially linear errors-in-variables model. The Annals of Statistics, 27: 1519–1535.

Liang H, Li R Z. 2009. Variable selection for partially linear models with measurement errors. Journal of the American Statistical Association, 104: 234–248.

Liang H, Liu X, Li R, et al. 2010. Estimation and testing for partially linear single-index models. The Annals of Statistics, 38: 3811–3836.

Liang H, Ren H. 2005. Generalized partially linear measurement error models. Journal of Computational and Graphical Statistics, 14: 237–250.

Liang H, Thurston S W, Ruppert D, et al. 2008. Additive partial linear models with measurement errors. Biometrika, 95(3): 667–678.

Liang H, Wang N. 2005. Partially linear single-index measurement error models. Statistica Sinica, 15: 99–116.

Lin D Y, Wei L J, Ying Z. 2002. Model-checking techniques based on cumulative residuals. Biometrics, 58(1): 1–12.

Lin X, Carroll R J. 2000. Nonparametric function estimation for clustered data when the predictor is measured without/with error. Journal of the American Statistical Association, 95: 520–534.

Liu Q. 2011. Estimation of the linear EV model with censored data. Journal of Statistical Planning and Inference, 141: 2463–2471.

Liu W, Wu L. 2008. A semiparametric nonlinear mixed-effects model with non-ignorable missing data and measurement errors for HIV viral data. Computational Statistics & Data Analysis, 53(1): 112–122.

Liu X, Wang L, Liang H. 2011. Estimation and variable selection for semiparametric additive partial linear models. Statistica Sinica, 21: 1225–1248.

Lu Y Q. 2008. Generalized partially linear varying-coefficient models. Journal of Statistical Planning and Inference, 138: 901–914.

Lu Y Q, Mao S S. 2004. Local asymptotics for B-spline estimators of the varying coefficient model. Communications in Statistics – Theory and Methods, 33: 1119–1138.

Lu Y Q, Zhang R Q, Zhu L P. 2008. Penalized spline estimation for varying- coefficient models. Communications in Statistics – Theory and Methods, 37: 2249–2261.

Lue H H. 2004. Principal Hessian directions for regression with measurement error. Biometrika, 91: 409–423.

Lue H H. 2008. Sliced average variance estimation for censored data. Communications in Statistics – Theory and Methods, 37: 3276–3286.

Lue H H. 2010. On principal Hessian directions for multivariate response regressions. Computational Statistics, 25: 619–632.

Luo R, Wang H, Tsai C L. 2009. Contour projected dimension reduction. The Annals of Statistics, 37: 3743–3778.

Ma S J, Liang H, Tsai C L. 2014. Partially linear single index models for repeated measurements. Journal of Multivariate Analysis, 130: 354–375.

Ma Y, Carroll R J. 2006. Locally efficient estimators for semiparametric models with measurement error. Journal of the American Statistical Association, 101: 1465–1474.

Ma Y, Yin G. 2011. Censored quantile regression with covariate measurement errors. Statistica Sinica, 21: 949–971.

Ma Y Y, Chiou J M, Wang N. 2006. Efficient semiparametric estimator for heteroscedastic partially linear models. Biometrika, 93(1): 75–84.

Ma Y Y, Li R Z. 2010. Variable selection in measurement error models. Bernoulli, 16: 274–300.

Mammen E. 1993. Bootstrap and wild bootstrap for high-dimensional linear models. The Annals of Statistics, 21(1): 255–285.

Mammen E, van de Geer S. 1997. Penalized estimation in partial linear models. The Annals of Statistics, 25: 1014–1035.

McCullagh P, Nelder J A. 1990. Generalized Linear Models. London: Chapmen & Hall/CRC.

Nadaraya E A. 1964. On estimating regression. Theory of Probability and Its Application, 9(1): 141–142.

Nakamura T. 1990. Corrected score functions for errors-in-variables models: methodology and application to generalized linear models. Biometrika, 77: 127–137.

Nguyen D V, Sentürk D. 2008. Multicovariate-adjusted regression models. Journal of Statistical Computation and Simulation, 78: 813–827.

Ni X, Zhang H, Zhang D. 2009. Automatic model selection for partially linear models. Journal of Multivariate Analysis, 100: 2100–2111.

Owen A B. 1988. Empirical likelihood ratio confidence intervals for a single function. Biometrika, 75(2): 237–249.

Owen A B. 1990. Empirical likelihood ratio confidence regions. The Annals of Statistics, 18: 90–120.

Owen A B. 1991. Empirical likelihood for linear models. The Annals of Statistics, 19(4): 1725–1747.

Owen A B. 1992. Empirical likelihood and generalized projection pursuit. Technical Report 393, Department of Statistics, Standford University.

Owen A B. 2001. Empirical Likelihood. New York: Chapman & Hall/CRC.

Pang Z, Xue L G. 2012. Estimation for the single-index models with random effects. Computational Statistics & Data Analysis, 56: 1837–1853.

Pepe M S. 1992. Inference using surrogate outcome data and a validation sample. Biometrika, 79: 355–365.

Poon W Y, Wang H B. 2014. Multivariate partially linear single-index models: Bayesian analysis. Journal of Nonparametric Statistics, 26: 755–768

Qin G S, Jing B Y. 2001. Empirical likelihood for censored linear regression. Scandinavian Journal of Statistics, 28: 661–673.

Qin J, Lawless J F. 1994. Empirical likelihood and general estimating equations. The Annals of Statistics, 22: 300–325.

Qu A, Li R. 2006. Quadratic inference functions for varying-coefficient models with longitudinal data. Biometrics, 62: 379–391.

Rao J N K, Scott A J. 1981. The analysis of categorical data from complex sample surveys:chi-squared tests for goodness of fit and independence in two-way tables. Journal of the American Statistical Association, 76: 221–230.

Rice J. 1986. Convergence rates for partial linear models. Statistics & Probability Letters, 4: 203–208.

Robinson P M. 1988. Root-n-consistent semiparametric regression. Econometrica, 56: 931–954.

Ruppert D, Sheather S J, Wand M P. 1995. An effective bandwidth selector for local least squares regression. Journal of the American Statistical Association, 90: 1257–1270.

Ruppert D, Wand M P, Carroll R J. 2003. Semiparametric Regression. New York: Cambridge University Press.

Schennach S M. 2004. Estimation of nonlinear models with measurement error. Econometrica, 72: 33–75.

Schennach S M. 2007. Instrumental variable estimation of nonlinear errors-in-variables models. Econometrica, 75: 201–239.

Schott J R. 1994. Determining the dimensionality in sliced inverse regression. Journal of the American Statistical Association, 89: 141–148.

Schwarz G. 1978. Estimating the dimension of a model. The Annals of Statistics, 6: 461–464.

Seber G A F, Wild C J. 1989. Nonlinear Regression. New York: John Wiley & Sons.

Sentürk D. 2006. Covariate-adjusted varying coefficient models. Biostatistics, 7: 235–251.

Sentürk D, Müller H G. 2005a. Covariate-adjusted regression. Biometrika, 92: 75–89.

Sentürk D, Müller H G. 2005b. Covariate adjusted correlation analysis via varying coeffi-

cient models. Scandinavian Journal of Statistics, 32: 365–383.

Sentürk D, Müller H G. 2006. Inference for covariate-adjusted regression via varying coefficient models. The Annals of Statistics, 34: 654–679.

Sentürk D, Müller H G. 2009. Covariate-adjusted generalized linear models. Biometrika, 96(2): 357–370.

Sentürk D, Nguyen D V. 2006. Estimation in covariate-adjusted regression. Computational Statistics & Data Analysis, 20: 3294–3310.

Sepanski J H, Lee L F. 1995. Semiparametric estimation of nonlinear error-in-variables models with validation study. Journal of Nonparametric Statistics, 4: 365–394.

Severini T A, Staniswalis J G. 1994. Quasilikelihood estimation in semiparametric models. Journal of American Statistical Association, 89: 501–511.

Shao F, Li J, Ma S, et al. 2014. Semiparametric varying-coefficient model for interval censored data with a cured proportion. Statistics in Medicine, 33(10): 1700–1712.

Shao Y W, Cook R D, Weisberg S. 2007. Marginal tests with sliced average variance estimation. Biometrika, 94: 285–296.

Sheil J, O'Muircheartaigh I. 1977. Algorithm AS106: The distribution of non-negative quadratic forms in normal variables. Applied Statistics, 26: 92–98.

Shi J, Lau T S. 2000. Empirical likelihood for partially linear models. Journal of Multivariate Analysis, 72(1): 132–148.

Shi P D. 1992. M estimation for partly linear models. Beijing: Ph.D Thesis, Institute of Systems Science, Chinese Academy of Sciences.

Silverman B W. 1986. Density Estimation for Statistics and Data Analysis. Monographs on Statistics and Applied Probability, London: Chapmen & Hall/CRC.

Sinha S, Ma Y. 2014. Semiparametric analysis of linear transformation models with covariate measurement errors. Biometrics, 70: 21–32.

Speckman P. 1988. Kernel smoothing in partial linear models. Journal of the Royal Statistical Society, Series B, 50: 413–436.

Staudenmayer J, Ruppert D. 2004. Local polynomial regression and simulation-extrapolation. Journal of the Royal Statistical Society, Series B (Statistical Methodology), 66(1): 17–30.

Stefanski L A, Bay J M. 1996. Simulation extrapolation deconvolution of finite population cumulative distribution function estimators. Biometrika, 83: 407–417.

Stefanski L A, Carroll R. 1987. Conditional scores and optimal scores for generalized linear measurement-error models. Biometrika, 74: 703–716.

Stefanski L A, Carroll R. 1989. Unbiased estimation of a nonlinear function of a normal mean with application to measurement error models. Communications in Statistics-Theory and Methods, 18: 4335–4358.

Stefanski L, Carroll R J. 1990. Deconvoluting kernel density estimators. Statistics, 21:

169–184.

Stein C. 1981. Estimation of the mean of a multivariate normal distribution. The Annals of Statistics, 9: 1135–1151.

Stone C J. 1982. Optimal global rates of convergence for nonparametric regression. The Annals of Statistics, 10: 1348–1360.

Stone C J. 1985. Additive regression and other nonparametric models. The Annals of Statistics, 13: 689–705.

Stone C J. 1986. The dimensionality reduction principle for generalized additive models. The Annals of Statistics, 14: 590–606.

Stute W, González Manteiga W, Presedo Quindimil M. 1998a. Bootstrap approximations in model checks for regression. Journal of American Statistical Association, 93: 141–149.

Stute W, Thies S, Zhu L X. 1998b. Model checks for regression: an innovation process approach. The Annals of Statistics, 26: 1916–1934.

Stute W, Xue L G, Zhu L X. 2007. Empirical likelihood inference in nonlinear errors-in-covariables models with validation data. Journal of the American Statistical Association, 102: 332–346.

Stute W, Zhu L X. 2005. Nonparametric checks for single-index models. The Annals of Statistics, 33: 1048–1083.

Suh E Y, Schafer D W. 2002. Semiparametric maximum likelihood for nonlinear regression with measurement errors. Biometrics, 58: 448–453.

Tang Q G, Cheng L S. 2008. M-estimation and B-spline approximation for varying coefficient models with longitudinal data. Journal of Nonparametric Statistics, 20: 611–625.

Tang Q G, Wang J D. 2005a. One-step estimation for varying coefficient models. Science in China, Series A-Mathematics, 48(2): 198–213.

Tang Q G, Wang J D. 2005b. L_1-estimation for varying coefficient models. Statistics, 39(5): 389–404.

Tang Y, Wang H, Zhu Z, et al. 2012. A unified variable selection approach for varying coefficient models. Statistica Sinica, 22: 601–628.

Taupin L. 2001. Semi-parametric estimation in the nonlinear structural errors-in-variables model. The Annals of Statistics, 29(1): 66–93.

Tibshirani R. 1996. Regression shrinkage and selection via the Lasso. Journal of the Royal Statistical Society, Series B, 58: 267–288.

Wahba G. 1977. A survey of some smoothing problems and the method of generalized cross-validation for solving them // Krisnaiah, P. R. Ed., Applications of Statistics. North Holland, Amsterdam, 507–523.

Wand M P, Jones M C. 1995. Kernel Smoothing. London: Chapmen & Hall/CRC.

Wang G N, Wang L. 2015. Spline estimation and variable selection for single-index prediction models with diverging number of index parameters. Journal of Statistical Planning

and Inference, 162: 1–19.

Wang H, Li B, Leng C. 2009a. Shrinkage tuning parameter selection with a diverging number of parameters. Journal of Royal Statistical Society, Series B, 71: 671–683.

Wang H, Li R, Tsai C L. 2007. Tuning parameter selectors for the smoothly clipped absolute deviation method. Biometrika, 94: 553–568.

Wang H, Xia Y. 2008. Sliced regression for dimension reduction. Journal of the American Statistical Association, 103: 811–821.

Wang H J, Zhu Z, Zhou J. 2009b. Quantile regression in partially linear varying coefficient models. The Annals of Statistics, 37: 3841–3866.

Wang H S. 2009. Rank reducible varying coefficient model. Journal of Statistical Planning and Inference, 139: 999–1011.

Wang H S, Xia Y C. 2009. Shrinkage estimation of the varying coefficient model. Journal of the American Statistical Association, 104: 747–757.

Wang J L, Xue L G, Zhu L X, et al. 2010. Estimation for a partial-linear single-index model. The Annals of Statistics, 38: 246–274.

Wang L. 2004. Estimation of nonlinear models with Berkson measurement errors. The Annals of Statistics, 32(6): 2559–2579.

Wang L, Hsiao C. 2011. Method of moments estimation and identifiability of semiparametric nonlinear errors-in-variables models. Journal of Econometrics, 165(1): 30–44.

Wang L, Li H, Huang J Z. 2008. Variable selection in nonparametric varying-coefficient models for analysis of repeated measurements. Journal of the American Statistical Association, 103(484): 1556–1569.

Wang L, Liu X, Liang H, et al. 2011. Estimation and variable selection for generalized additive partial linear models. The Annals of Statistics, 39: 1827–1851.

Wang Q, Yin X. 2008. A nonlinear multi-dimensional variable selection method for high dimensional data: sparse MAVE. Computational Statistics & Data Analysis, 52: 4512–4520.

Wang Q H. 1999. Estimation of partial linear error-in-variables models with validation data. Journal of Multivariate Analysis, 69: 30–64.

Wang Q H. 2003. Dimension reduction in partly linear error-in-response models with validation data. Journal of Multivariate Analysis, 85: 234–252.

Wang Q H, Jing B Y. 1999. Empirical likelihood for partial linear models with fixed designs. Statistics & Probability Letters, 41(4): 425–433.

Wang Q H, Rao J N K. 2002a. Empirical likelihood-based inference in linear errors-in-covariables models with validation data. Biometrik, 89(2): 345–357.

Wang Q H, Rao J N K. 2002b. Empirical likelihood-based inference in linear models with missing data. Scandinavian Journal of Statistics, 29: 563–576.

Wang Q H, Sun Z H. 2007. Estimation in partially linear models with missing responses

at random. Journal of Multivariate Analysis, 98: 1470–1493.

Wang Q H, Zhang R Q. 2009. Statistical estimation in varying coefficient models with surrogate data and validation sampling. Journal of Multivariate Analysis, 100: 2389–2405.

Wang Q H, Zhu L X. 2001. Estimation in partly linear errors-in-variable models with censored data. Communications in Statistics – Theory and Methods, 30: 41–54.

Wang X G, Shi X Y. 2014. Robust estimation for survival partially linear single-index models. Computational Statistics & Data Analysis, 80: 140–152.

Wang X L, Li G R, Lin L. 2011. Empirical likelihood inference for semi-parametric varying-coefficient partially linear EV models. Metrika, 73: 171–185.

Watson G S. 1964. Smooth regression analysis. Sankhya: The Indian Journal of Statistics, Series A, 26(4): 359–372.

Wedderburn R W M. 1974. Quasi-likelihood functions, generalized linear models and the Gauss-Newton method. Biometrika, 61: 439–447.

Wei C H. 2011. Estimation in varying-coefficient errors-in-variables models with missing response variables. Communications in Statistics—Simulation and Computation, 40(3): 383–393.

Wei C H. 2012. Statistical inference for restricted partially linear varying coefficient errors-in-variables models. Journal of Statistical Planning and Inference, 142: 2464–2472.

Wei F, Huang J, Li H. 2011. Variable selection and estimation in high-dimensional varying coefficient models. Statistica Sinica, 21(4): 1515–1540.

Wittes J, Lakators E, Protstfield J. 1989. Surrogate endpoints in clinical trials: cardiovascular diseases. Statistics in Medicine, 8: 415–425.

Wolter K M, Fuller W A. 1982. Estimation of nonlinear errors-in-variables models. The Annals of Statistics, 10(2): 539–548.

Wong H, Liu F, Chen M, et al. 2009. Empirical likelihood based diagnostics for heteroscedasticity in partially linear errors-in-variables models. Journal of Statistical Planning and Inference, 139: 916–929.

Wu C O, Chiang C T. 2000. Kernel smoothing on varying coefficient models with longitudinal dependent variable. Statistica Sinica, 10: 433–456.

Wu C O, Chiang C T, Hoover D R. 1998. Asymptotic confidence regions for kernel smoothing of a varying-coefficient model with longitudinal data. Journal of the American Statistical Association, 93: 1388–1403.

Xia Y C. 2007. A constructive approach to the estimation of dimension reduction directions. The Annals of Statistics, 35: 2654–2690.

Xia Y C, Hädle W. 2006. Semi-parametric estimation of partially linear single-index models. Journal of Multivariate Analysis. 97: 1162–1184.

Xia Y C, Tong H, Li W K, et al. 2002. An adaptive estimation of dimension reduction

space. Journal of the Royal Statistical Society, Series B, 64: 363–410.

Xia Y C, Zhang W, Tong H. 2004. Efficient estimation for semivarying-coefficient models. Biometrika, 91(3): 661–681.

Xie H, Huang J. 2009. SCAD-Penalized regression in high-dimensional partially linear models. The Annals of Statistics, 37: 673–696.

Xu J, Leng C, Ying Z. 2010. Rank-based variable selection in the accelerated failure time model. Statistics and Computing, 20: 165–176.

Xue L G. 2006. Empirical likelihood inference in nonlinear semiparametric EV models with validation data. Acta Mathematica Sinica, Chinese Series, 49(1): 145–154.

Xue L G. 2009a. Empirical likelihood confidence intervals for response mean with data missing at random. Scandinavian Journal of Statistics, 36: 671–685.

Xue L G. 2009b. Empirical likelihood for linear models with missing responses. Journal of Multivariate Analysis, 100: 1353–1366.

Xue L G, Zhu L X. 2007. Empirical likelihood for a varying coefficient model with longitudinal data. Journal of the American Statistical Association, 102: 642–654.

Xue L G, Zhu L X. 2010. Empirical Likelihood in Nonparametric and Semiparametric Models. Beijing: Science Press.

Yan L, Chen X. 2014. Empirical likelihood for partly linear models with errors in all variables. Journal of Multivariate Analysis, 130: 275–288.

Yang H, Li T T. 2010. Empirical likelihood for semiparametric varying coefficient partially linear models with longitudinal data. Statistics & Probability Letters, 80(2): 111–121.

Yang S G, Xue L G, Li G R. 2014. Simultaneous confidence bands for single-index random effects models with longitudinal data. Statistics & Probability Letters, 85: 6–14.

Yang Y P, Li G R, Tong T J. 2015. Corrected empirical likelihood for a class of generalized linear measurement error models. Science China Mathematics, 58(7): 1523–1536.

Ye Z, Weiss. 2003. Using the bootstrap to select one of a new class of dimension reduction methods. Journal of the American Statistical Association, 98: 968–979.

Yi G Y, Liu W, Wu L. 2011. Simultaneous inference and bias analysis for longitudinal data with covariate measurement error and missing responses. Biometrics, 67: 67–75.

Yi G Y, Ma Y Y, Carroll R J. 2012. A functional generalized method of moments approach for longitudinal studies with missing responses and covariate measurement error. Journal of the American Statistical Association, 99: 151–165.

You J H, Chen G M. 2006. Estimation of a semiparametric varying-coefficient partially linear errors-in-variables model. Journal of Multivariate Analysis, 97: 324–341.

You J H, Chen G M, Zhou Y. 2006. Block empirical likelihood for longitudinal partially linear regression models. The Canadian Journal of Statistics, 34(1): 79–96.

You J H, Zhou Y. 2006. Empirical Likelihood for Semi-parametric Varying-Coefficient Partially Linear Model. Statistics & Probability Letters, 76: 412–422.

Zeng P. 2008. Determining the dimension of the central subspace and central mean subspace. Biometrika, 95: 469–479.

Zhang J, Feng Z H, Xu P R. 2015a. Estimating the conditional single-index error distribution with a partial linear mean regression. Test, 24: 61–83.

Zhang J, Gai Y, Wu P. 2013a. Estimation in linear regression models with measurement errors subject to single-indexed distortion. Computational Statistics & Data Analysis, 59: 103–120.

Zhang J, Li G R, Feng Z H. 2015b. Checking the adequacy for a distortion errors-in-variables parametric regression model. Computational Statistics & Data Analysis, 83: 52–64.

Zhang J, Wang T, Liang H, et al. 2012c. A dimension reduction based approach for estimation and variable selection in partially linear single-index models with high dimensional covariates. Electronic Journal of Statistics, 6: 2235–2273.

Zhang J, Wang X G, Yu Y, et al. 2014a. Estimation and variable selection in partial linear single index models with error-prone linear covariates. Statistics: A Journal of Theoretical and Applied Statistics, 48: 1048–1070.

Zhang J, Yu Y, Zhou B, et al. 2014b. Nonlinear measurement errors models subject to additive distortion, Journal of Statistical Planning and Inference, 150: 49–65.

Zhang J, Yu Y, Zhu L X, et al. 2013b. Partial linear single index models with distortion measurement errors. Annals of the Institute of Statistical Mathematics, 65: 237–267.

Zhang J, Zhu L X, Liang H. 2012a. Nonlinear models with measurement errors subject to single-indexed distortion. Journal of Multivariate Analysis, 112: 1–23.

Zhang J, Zhu L P, Zhu L X. 2012b. On a dimension reduction regression with covariate adjustment. Journal of Multivariate Analysis, 104: 39–55.

Zhang J, Zhu L P, Zhu L X. 2014c. Surrogate dimension reduction for measurement error regressions. Statistica Sinica, 24: 1341–1363.

Zhang R, Zhao W, Liu J. 2013c. Robust estimation and variable selection for semiparametric partially linear varying coefficient model based on modal regression. Journal of Nonparametric Statistics, 25(2): 523–544.

Zhang X, Liang H. 2011. Focused information criterion and model averaging for generalized additive partial linear models. The Annals of Statistics, 39: 194–200.

Zhang W W, Li G R, Xue L G. 2011. Profile inference on partially linear varying-coefficient errors-in-variables models under restricted condition. Computational Statistics & Data Analysis, 55: 3027–3040.

Zhang W Y, Lee S Y. 2000. Variable bandwidth selection in vary-coefficient models. Journal of Multivariate Analysis, 74: 116–134.

Zhang W Y, Lee S Y, Song X Y. 2002. Local polynomial fitting in semivarying coefficient model. Journal of Multivariate Analysis, 82: 166–188.

Zhang W Y, Peng H. 2010. Simultaneous confidence band and hypothesis test in generalised varying-coefficient models. Journal of Multivariate Analysis, 101(7): 1656–1680.

Zhang X, Liang, H. 2011. Focused information criterion and model averaging for generalized additive partial linear models. The Annals of Statistics, 39: 194–200.

Zhao L C, Krishnaiah P R, Bai Z D. 1986a. On detection of the number of signals in presence of white noise. Journal of Multivariate Analysis, 20: 1–25.

Zhao L C, Krishnaiah P R, Bai Z D. 1986b. On detection of the number of signals when the noise covariance matrix is arbitrary. Journal of Multivariate Analysis, 20: 26–49.

Zhao P X, Xue L G. 2009. Variable selection for semiparametric varying coefficient partially linear models. Statistics & Probability Letters, 79: 2148–2157.

Zhao P X, Xue L G. 2010. Variable selection for semiparametric varying coefficient partially linear errors-in-variables models. Journal of Multivariate Analysis, 101: 1872–1883.

Zhao P X, Xue L G. 2011. Variable selection for varying coefficient models with measurement errors. Metrika, 74: 231–245.

Zhou X, You J H. 2004. Wavelet estimation in varying-coefficient partially linear regression models. Statistics & Probability Letters, 68(1): 91–104.

Zhou Y, Liang H. 2009. Statistical inference for semiparametric varying-coefficient partially linear models with error-prone linear covariates. The Annals of Statistics, 37: 427–458.

Zhu L P, Zhu L X. 2009a. On kernel method for sliced average variance estimation. Journal of Multivariate Analysis, 21: 851–861.

Zhu L P, Zhu L X. 2009b. Nonconcave penalized inverse regression in single-index models with high dimensional predictors. Journal of Multivariate Analysis, 100: 862–875.

Zhu L P, Zhu L X, Feng, Z H. 2010. Dimension reduction in regressions through cumulative slicing estimation. Journal of the American Statistical Association, 105: 1455–1466.

Zhu L X, Cui H J. 2003. A semi-parametric regression model with errors in variables. Scandinavian Journal of Statistics, 30: 429–442.

Zhu L X, Fang K T. 1996. Asymptotics for kernel estimate of sliced inverse regression. The Annals of Statistics, 24: 1053–1068.

Zhu L X, Miao B, Peng H. 2006. On sliced inverse regression with large dimensional covariates. Journal of the American Statistical Association, 101: 630–643.

Zhu L X, Ng K W. 1995. Asymptotics for sliced inverse regression. Statistica Sinica, 5: 727–736.

Zhu L X, Xue L G. 2006. Empirical likelihood confidence regions in a partially linear single-index model. Journal of the Royal Statistical Society, Series B, 68: 549–570.

Zhu X H, Guo X, Lin L, et al. 2015. Heteroscedasticity checks for single index models. Journal of Multivariate Analysis, 136: 41–55.

Zou H. 2006. The adaptive Lasso and its oracle properties. Journal of the American Statistical Association, 101: 1418–1429.

Zou H, Hastie T. 2005. Regularization and variable selection via the elastic net. Journal of the Royal Statistical Society, Series B, 67: 301–320.

Zou H, Li R Z. 2008. One-step sparse estimates in nonconcave penalized likelihood models (with discussion). The Annals of Statistics, 36(4): 1509–1533.

索　引

(按拼音字母排序)

G

H

J

其他

《现代数学基础丛书》已出版书目

(按出版时间排序)

1 数理逻辑基础(上册) 1981.1 胡世华 陆钟万 著
2 紧黎曼曲面引论 1981.3 伍鸿熙 吕以辇 陈志华 著
3 组合论(上册) 1981.10 柯 召 魏万迪 著
4 数理统计引论 1981.11 陈希孺 著
5 多元统计分析引论 1982.6 张尧庭 方开泰 著
6 概率论基础 1982.8 严士健 王隽骧 刘秀芳 著
7 数理逻辑基础(下册) 1982. 8 胡世华 陆钟万 著
8 有限群构造(上册) 1982.11 张远达 著
9 有限群构造(下册) 1982.12 张远达 著
10 环与代数 1983.3 刘绍学 著
11 测度论基础 1983.9 朱成熹 著
12 分析概率论 1984.4 胡迪鹤 著
13 巴拿赫空间引论 1984.8 定光桂 著
14 微分方程定性理论 1985.5 张芷芬 丁同仁 黄文灶 董镇喜 著
15 傅里叶积分算子理论及其应用 1985.9 仇庆久等 编
16 辛几何引论 1986.3 J.柯歇尔 邹异明 著
17 概率论基础和随机过程 1986.6 王寿仁 著
18 算子代数 1986.6 李炳仁 著
19 线性偏微分算子引论(上册) 1986.8 齐民友 著
20 实用微分几何引论 1986.11 苏步青等 著
21 微分动力系统原理 1987.2 张筑生 著
22 线性代数群表示导论(上册) 1987.2 曹锡华等 著
23 模型论基础 1987.8 王世强 著
24 递归论 1987.11 莫绍揆 著
25 有限群导引(上册) 1987.12 徐明曜 著
26 组合论(下册) 1987.12 柯 召 魏万迪 著
27 拟共形映射及其在黎曼曲面论中的应用 1988.1 李 忠 著
28 代数体函数与常微分方程 1988.2 何育赞 著
29 同调代数 1988.2 周伯壎 著

30　近代调和分析方法及其应用　1988.6　韩永生　著
31　带有时滞的动力系统的稳定性　1989.10　秦元勋等　编著
32　代数拓扑与示性类　1989.11　马德森著　吴英青　段海鲍译
33　非线性发展方程　1989.12　李大潜　陈韵梅　著
34　反应扩散方程引论　1990.2　叶其孝等　著
35　仿微分算子引论　1990.2　陈恕行等　编
36　公理集合论导引　1991.1　张锦文　著
37　解析数论基础　1991.2　潘承洞等　著
38　拓扑群引论　1991.3　黎景辉　冯绪宁　著
39　二阶椭圆型方程与椭圆型方程组　1991.4　陈亚浙　吴兰成　著
40　黎曼曲面　1991.4　吕以辇　张学莲　著
41　线性偏微分算子引论(下册)　1992.1　齐民友　许超江　编著
42　复变函数逼近论　1992.3　沈燮昌　著
43　Banach 代数　1992.11　李炳仁　著
44　随机点过程及其应用　1992.12　邓永录等　著
45　丢番图逼近引论　1993.4　朱尧辰等　著
46　线性微分方程的非线性扰动　1994.2　徐登洲　马如云　著
47　广义哈密顿系统理论及其应用　1994.12　李继彬　赵晓华　刘正荣　著
48　线性整数规划的数学基础　1995.2　马仲蕃　著
49　单复变函数论中的几个论题　1995.8　庄圻泰　著
50　复解析动力系统　1995.10　吕以辇　著
51　组合矩阵论　1996.3　柳柏濂　著
52　Banach 空间中的非线性逼近理论　1997.5　徐士英　李　冲　杨文善　著
53　有限典型群子空间轨道生成的格　1997.6　万哲先　霍元极　著
54　实分析导论　1998.2　丁传松等　著
55　对称性分岔理论基础　1998.3　唐　云　著
56　Gel'fond-Baker 方法在丢番图方程中的应用　1998.10　乐茂华　著
57　半群的 S-系理论　1999.2　刘仲奎　著
58　有限群导引(下册)　1999.5　徐明曜等　著
59　随机模型的密度演化方法　1999.6　史定华　著
60　非线性偏微分复方程　1999.6　闻国椿　著
61　复合算子理论　1999.8　徐宪民　著
62　离散鞅及其应用　1999.9　史及民　编著
63　调和分析及其在偏微分方程中的应用　1999.10　苗长兴　著

64　惯性流形与近似惯性流形　2000.1　戴正德　郭柏灵　著
65　数学规划导论　2000.6　徐增堃　著
66　拓扑空间中的反例　2000.6　汪　林　杨富春　编著
67　拓扑空间论　2000.7　高国士　著
68　非经典数理逻辑与近似推理　2000.9　王国俊　著
69　序半群引论　2001.1　谢祥云　著
70　动力系统的定性与分支理论　2001.2　罗定军　张　祥　董梅芳　编著
71　随机分析学基础(第二版)　2001.3　黄志远　著
72　非线性动力系统分析引论　2001.9　盛昭瀚　马军海　著
73　高斯过程的样本轨道性质　2001.11　林正炎　陆传荣　张立新　著
74　数组合地图论　2001.11　刘彦佩　著
75　光滑映射的奇点理论　2002.1　李养成　著
76　动力系统的周期解与分支理论　2002.4　韩茂安　著
77　神经动力学模型方法和应用　2002.4　阮炯　顾凡及　蔡志杰　编著
78　同调论——代数拓扑之一　2002.7　沈信耀　著
79　金兹堡-朗道方程　2002.8　郭柏灵等　著
80　排队论基础　2002.10　孙荣恒　李建平　著
81　算子代数上线性映射引论　2002.12　侯晋川　崔建莲　著
82　微分方法中的变分方法　2003.2　陆文端　著
83　周期小波及其应用　2003.3　彭思龙　李登峰　谌秋辉　著
84　集值分析　2003.8　李　雷　吴从炘　著
85　数理逻辑引论与归结原理　2003.8　王国俊　著
86　强偏差定理与分析方法　2003.8　刘　文　著
87　椭圆与抛物型方程引论　2003.9　伍卓群　尹景学　王春朋　著
88　有限典型群子空间轨道生成的格(第二版)　2003.10　万哲先　霍元极　著
89　调和分析及其在偏微分方程中的应用(第二版)　2004.3　苗长兴　著
90　稳定性和单纯性理论　2004.6　史念东　著
91　发展方程数值计算方法　2004.6　黄明游　编著
92　传染病动力学的数学建模与研究　2004.8　马知恩　周义仓　王稳地　靳　祯　著
93　模李超代数　2004.9　张永正　刘文德　著
94　巴拿赫空间中算子广义逆理论及其应用　2005.1　王玉文　著
95　巴拿赫空间结构和算子理想　2005.3　钟怀杰　著
96　脉冲微分系统引论　2005.3　傅希林　闫宝强　刘衍胜　著
97　代数学中的 Frobenius 结构　2005.7　汪明义　著

98 生存数据统计分析 2005.12 王启华 著

99 数理逻辑引论与归结原理(第二版) 2006.3 王国俊 著

100 数据包络分析 2006.3 魏权龄 著

101 代数群引论 2006.9 黎景辉 陈志杰 赵春来 著

102 矩阵结合方案 2006.9 王仰贤 霍元极 麻常利 著

103 椭圆曲线公钥密码导引 2006.10 祝跃飞 张亚娟 著

104 椭圆与超椭圆曲线公钥密码的理论与实现 2006.12 王学理 裴定一 著

105 散乱数据拟合的模型方法和理论 2007.1 吴宗敏 著

106 非线性演化方程的稳定性与分歧 2007.4 马 天 汪守宏 著

107 正规族理论及其应用 2007.4 顾永兴 庞学诚 方明亮 著

108 组合网络理论 2007.5 徐俊明 著

109 矩阵的半张量积:理论与应用 2007.5 程代展 齐洪胜 著

110 鞅与 Banach 空间几何学 2007.5 刘培德 著

111 非线性常微分方程边值问题 2007.6 葛渭高 著

112 戴维-斯特瓦尔松方程 2007.5 戴正德 蒋慕蓉 李栋龙 著

113 广义哈密顿系统理论及其应用 2007.7 李继彬 赵晓华 刘正荣 著

114 Adams 谱序列和球面稳定同伦群 2007.7 林金坤 著

115 矩阵理论及其应用 2007.8 陈公宁 著

116 集值随机过程引论 2007.8 张文修 李寿梅 汪振鹏 高 勇 著

117 偏微分方程的调和分析方法 2008.1 苗长兴 张 波 著

118 拓扑动力系统概论 2008.1 叶向东 黄 文 邵 松 著

119 线性微分方程的非线性扰动(第二版) 2008.3 徐登洲 马如云 著

120 数组合地图论(第二版) 2008.3 刘彦佩 著

121 半群的 S-系理论(第二版) 2008.3 刘仲奎 乔虎生 著

122 巴拿赫空间引论(第二版) 2008.4 定光桂 著

123 拓扑空间论(第二版) 2008.4 高国士 著

124 非经典数理逻辑与近似推理(第二版) 2008.5 王国俊 著

125 非参数蒙特卡罗检验及其应用 2008.8 朱力行 许王莉 著

126 Camassa-Holm 方程 2008.8 郭柏灵 田立新 杨灵娥 殷朝阳 著

127 环与代数(第二版) 2009.1 刘绍学 郭晋云 朱 彬 韩 阳 著

128 泛函微分方程的相空间理论及应用 2009.4 王 克 范 猛 著

129 概率论基础(第二版) 2009.8 严士健 王隽骧 刘秀芳 著

130 自相似集的结构 2010.1 周作领 瞿成勤 朱智伟 著

131 现代统计研究基础 2010.3 王启华 史宁中 耿 直 主编

132　图的可嵌入性理论(第二版)　2010.3　刘彦佩　著
133　非线性波动方程的现代方法(第二版)　2010.4　苗长兴　著
134　算子代数与非交换 L_p 空间引论　2010.5　许全华　吐尔德别克　陈泽乾　著
135　非线性椭圆型方程　2010.7　王明新　著
136　流形拓扑学　2010.8　马　天　著
137　局部域上的调和分析与分形分析及其应用　2011.6　苏维宜　著
138　Zakharov 方程及其孤立波解　2011.6　郭柏灵　甘在会　张景军　著
139　反应扩散方程引论(第二版)　2011.9　叶其孝　李正元　王明新　吴雅萍　著
140　代数模型论引论　2011.10　史念东　著
141　拓扑动力系统——从拓扑方法到遍历理论方法　2011.12　周作领　尹建东　许绍元　著
142　Littlewood-Paley 理论及其在流体动力学方程中的应用　2012.3　苗长兴　吴家宏　章志飞　著
143　有约束条件的统计推断及其应用　2012.3　王金德　著
144　混沌、Mel'nikov 方法及新发展　2012.6　李继彬　陈凤娟　著
145　现代统计模型　2012.6　薛留根　著
146　金融数学引论　2012.7　严加安　著
147　零过多数据的统计分析及其应用　2013.1　解锋昌　韦博成　林金官　编著
148　分形分析引论　2013.6　胡家信　著
149　索伯列夫空间导论　2013.8　陈国旺　编著
150　广义估计方程估计方法　2013.8　周　勇　著
151　统计质量控制图理论与方法　2013.8　王兆军　邹长亮　李忠华　著
152　有限群初步　2014.1　徐明曜　著
153　拓扑群引论(第二版)　2014.3　黎景辉　冯绪宁　著
154　现代非参数统计　2015.1　薛留根　著
155　三角范畴与导出范畴　2015.5　章　璞　著
156　线性算子的谱分析(第二版)　2015.6　孙　炯　王　忠　王万义　编著
157　双周期弹性断裂理论　2015.6　李　星　路见可　著
158　电磁流体动力学方程与奇异摄动理论　2015.8　王　术　冯跃红　著
159　算法数论(第二版)　2015.9　裴定一　祝跃飞　编著
160　偏微分方程现代理论引论　2016.1　崔尚斌　著
161　有限集上的映射与动态过程——矩阵半张量积方法　2015.11　程代展　齐洪胜　贺风华　著
162　现代测量误差模型　2016.3　李高荣　张　君　冯三营　著